# Making Sense of Sensitivity: Extending Omitted Variable Bias

Carlos Cinelli * Chad Hazlett†

12th July 2018

Abstract

In this paper we extend the familiar "omitted variable bias" framework, creating a suite of tools for sensitivity analysis of regression coefficients and their standard errors to unobserved confounders that: (i) do not require assumptions about the functional form of the treatment assignment mechanism nor the distribution of the unobserved confounder(s); (ii) can be used to assess the sensitivity to multiple confounders, whether they influence the treatment or the outcome linearly or not; (iii) facilitate the use of expert knowledge to judge the plausibility of sensitivity parameters; and, (iv) can be easily and intuitively displayed, either in concise regression tables or more elaborate graphs. More precisely, we introduce two novel measures for communicating the sensitivity of regression results that can be used for routine reporting. The "robustness value" describes the association unobserved confounding would need to have with *both* the treatment *and* the outcome to change the research conclusions. The partial $R^2$ of the treatment with the outcome shows how strongly confounders explaining *all* of the outcome would have to be associated with the treatment to eliminate the estimated effect. Next, we provide intuitive graphical tools that allow researchers to make more elaborate arguments about the sensitivity of not only point estimates but also t-values (or p-values and confidence intervals). We also provide graphical tools for exploring extreme sensitivity scenarios in which *all* or much of the residual variance is assumed to be due to confounders. Finally, we note that a widespread informal "benchmarking" practice can be widely misleading, and introduce a novel alternative that allows researchers to formally bound the strength of unobserved confounders "as strong as" certain covariate(s) in terms of the explained variance of the treatment and/or the outcome. We illustrate these methods with a running example that estimates the effect of exposure to violence in western Sudan on attitudes toward peace.

*Graduate Student, Dept. of Statistics, University of California Los Angeles.
Email: carloscinelli@ucla.edu.

†Assistant Professor, Departments of Statistics and Political Science, University of California Los Angeles. Email: chazlett@ucla.edu URL: http://www.chadhazlett.com. We thank Michael Tzen, Brandon Stewart, Christopher M. Felton, Ian Lundberg, Kosuke Imai, Erin Hartman, Darin Christensen, and members of the Improving Design in Social Science workshop at UCLA for valuable comments and feedback. Thanks to Aaron Rudkin for assistance developing the forthcoming `R` package, `sensemakr`. Thanks to Fernando Mello for his examination of how many papers in Political Science journals have employed formal sensitivity analyses.

# Contents

# 1 Introduction

Observational research often seeks to estimate causal effects under a "no unobserved confounding" or "ignorability" (conditional on observables) assumption (see e.g. Rosenbaum and Rubin 1983b; Pearl 2009; Imbens and Rubin 2015). When making causal claims from observational data, investigators marshal what evidence they can to argue that their result is not due to confounding. In "natural" and "quasi"-experiments, this often includes a qualitative account for why the treatment assignment is "as-if" random conditional on a set of key characteristics (see e.g. Angrist and Pischke 2008; Dunning 2012). Investigators seeking to make causal claims from observational data are also instructed to show "balance tests" and "placebo tests". While, in some cases, null findings on these tests may be consistent with the claim of no unobserved confounders, they are certainly not dispositive: it is *unobserved* variables that we worry may be both "imbalanced" and related to the outcome in problematic ways. Fundamentally, causal inference always require assumptions that are unverifiable from the data (Pearl 2009).

Thus, in addition to balance and placebo tests, investigators are advised to conduct "sensitivity analyses" examining how fragile a result is against the possibility of unobserved confounding.[1] In general, such analyses entail two components: (1) describing the type of unobserved confounders—parameterized by their relation to the treatment assignment, the outcome, or both—that would substantively change our conclusions about the estimated causal effect, and (2) assisting the investigator in assessing the plausibility that such problematic confounding might exist, which necessarily depends upon the research design and expert knowledge regarding the data generating process. A variety of sensitivity analyses have been proposed, dating back to Cornfield et al. (1959), with more recent contributions including Rosenbaum and Rubin (1983a); Heckman et al. (1998); Robins (1999); Frank (2000); Rosenbaum (2002); Imbens (2003); Brumback et al. (2004); Hosman et al. (2010); Imai et al. (2010); Vanderweele and Arah (2011); Blackwell (2013); Frank et al. (2013); Dorie et al. (2016); Middleton et al. (2016) and Oster (2017). Yet, such sensitivity analyses remain underutilized.[2]

We argue that a number of factors contribute to this reluctant uptake. One is the complicated nature and strong assumptions many of these methods impose, often involving restrictions on or even a complete description of the nature of the confounder (see Section 6 for discussion). A second reason is that, while training, convention and convenience dictate that users routinely report "regression tables" (or perhaps coefficient plots) to convey the results of a regression, we lack readily available quantities that aid in understanding and communicating how sensitive our results are to potential unobserved confounding. Third, and most fundamentally, connecting the results of a formal sensitivity analysis to a cogent argument about what types of confounders may exist in one's research project is often difficult, particularly with research designs that do not hinge on a credible argument regarding the (conditionally) "ignorable", "exogeneous", or "as-if random" nature of the treatment. To complicate things, some of the solutions offered by the literature can lead users to erroneous conclusions.

In this paper we show how the familiar omitted variable bias (OVB) framework can be extended to address these challenges. We develop a suite of sensitivity analysis tools that do not require assumptions on the treatment assignment mechanism nor on the distribution of the unobserved

[1] Researchers may also wish to examine sensitivity to the choice of observed covariates, see Leamer (2016).

[2] In political science, out of 164 quantitative papers in the top three general interest publications (American Political Science Review, American Journal of Political Sciene, and Journal of Politics) for 2017, 64 papers clearly described a causal identification strategy other than a randomized experiment. Of these only 4 (6.25%) employed a formal sensitivity analyses beyond trying various specifications. In economics, Oster (2014) reports that most of non-experimental empirical papers utilized only informal robustness tests based on coefficient stability in the face of adding or dropping covariates. See also Chen and Pearl (2015).

confounder, and can be used to assess the sensitivity to multiple confounders, whether they influence the treatment and outcome linearly or not.

We first introduce two novel measures of the sensitivity of linear regression coefficients: (i) the "robustness value" (RV), which provides a convenient reference point to assess the overall robustness of a coefficient to unobserved confounding. If the confounders' association to the treatment and to the outcome (measured in terms of partial $R^2$) are *both* assumed to be less than the robustness value, then such confounders cannot "explain away" the observed effect. And, (ii) the proportion of variation in the outcome explained uniquely by the treatment, $R^2_{Y\sim D|\boldsymbol{X}}$, which reveals how strongly counfounders that explain 100% of the residual variance of the outcome would have to be associated with the treatment in order to eliminate the effect. To advance standard practice across a variety of disciplines, we propose routinely reporting the RV and $R^2_{Y\sim D|\boldsymbol{X}}$ in regression tables.

Next, we offer graphical tools that investigators can use to refine their sensitivity analyses. The first is close in spirit to the proposal of Imbens (2003)—a bivariate sensitivity contour plot, parameterizing the confounder in terms of partial $R^2$ values. However, contrary to Imbens' maximum likelihood approach, the OVB-based approach makes the underlying analysis simpler to understand, easier to compute, and more general. It side-steps assumptions about the distribution of the (possibly multiple, non-linear) confounders, and it easily extends contour plots to assess the sensitivity of t-values, p-values or confidence intervals. This enables users to examine the types of confounders that would alter their inferential conclusions, not just point estimates. The second is an "extreme-scenario" sensitivity plot, in which investigators make conservative assumptions about the portion of otherwise unexplainend variance in the outcome that is due to confounders. One can then see how strongly such confounders would need to be associated with the treatment to be problematic. In the "worst-case" of these scenarios, the investigator assumes *all* unexplained variation in the outcome may be due to a confounder.

Finally, we introduce a novel bounding procedure that aids researchers in judging which confounders are plausible or could be ruled out, using the observed data in combination with expert knowledge. While prior work (Imbens 2003; Hosman et al. 2010; Blackwell 2013; Dorie et al. 2016; Middleton et al. 2016) has suggested an informal practice of benchmarking the unobserved confounding by comparison to observables, we show that this practice can be widely misleading due to the effects of confounding itself, even if the confounder is assumed to be independent of the covariate(s) used for benchmarking. Instead, our approach formally bounds the strength of unobserved confounding with the same strength (or a multiple thereof) as a chosen observable or group of observables. These bounds are tight and may be especially useful when investigators can credibly argue to have measured the most important determinants of the treatment assignment or of the outcome.

In what follows, Section 2 describes the running example—a study of the effect of violence on attitudes toward peace in Darfur, Sudan—that will be used to illustrate the tools throughout the text. Section 3 introduces the traditional OVB framework, how it can be used for a first approach to sensitivity analysis, and some of its shortcomings. Next, Section 4 shows how to extend the traditional OVB with the partial $R^2$ parameterization and Section 5 demonstrates how these results can be used to enrich sensitivity analysis in practice. We end the paper by speaking to how our proposal compares to existing ones, how it can help with the dissemination of sensitivity analysis, and highlighting important caveats when interpreting sensitivity results. Open-source software for `R` implements the methods presented here[3].

[3] Forthcoming.

## 2 Running example

In this section we briefly introduce the applied example used throughout the paper.[4] This serves as a background to illustrate how the tools developed here can be applied to address problems that commonly arise in observational research. Another point we emphasize is that the information produced by a sensitivity analysis is useful to the extent that researchers can wield expert knowledge to rule out the types of confounders shown to be problematic. Thus, a real world example helps to illustrate how contextual knowledge could be employed.

### 2.1 Exposure to violence in Darfur

In Sudan's western region of Darfur, a horrific campaign of violence against civilians began in 2003, sustaining high levels of violence through 2004, and killing an estimated 200,000 (Flint and de Waal 2008). It was deemed genocide by then Secretary of State Colin Powell, and has resulted in indictments of alleged genocide, war crimes, and crimes against humanity in the International Criminal Court.

In the current case, we are interested in learning how exposure to this violence changed individual attitudes towards peace, and in particular how being physically harmed during attacks on one's villages influences attitudes. Clearly we cannot randomize who is exposed to such violence. However, the means by which violence was distributed provide a tragic natural experiment. Violence against civilians in villages during this time included both aerial bombardments by government aircraft, and attacks by a pro-government militia called the *Janjaweed*. In short, while some villages were singled out for more or less violence, within a given village violence was arguably indiscriminate. Such arguments are supported by reports such as

> The government came with antonovs, and targeted everything that moved. They made no distinction between the civilians and rebel groups. If it moved, it was bombed. It is the same thing, whether there are rebel groups (present) or not...The government bombs from the sky and the *Janjaweed* sweeps through and burns everything and loots the animals and spoils everything that they cannot take (Human Rights Watch 2004)

Furthermore, one may argue for the indiscriminacy-within-village claim on the basis that the violence promoted by the government was mainly used to drive people out rather than target individuals. The bombing was crude, could not be targeted within village, and the attackers had almost no information about who they would target within a village. There is one major exception to this: while both men and women were often injured or killed, women were also targeted for widespread sexual assault and rape by the *Janjaweed*. For this reason, gender may also be a critical variable to condition on. With this in mind, we may estimate the linear model,

$$\text{PeaceIndex} = \hat{\tau}_{\text{res}}\text{DirectHarm} + \hat{\beta}_{f,\text{res}}\text{Female} + \text{Village}\hat{\beta}_{v,\text{res}} + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\varepsilon}_{\text{res}} \tag{1}$$

where *PeaceIndex* is an index measuring individual attitudes towards peace, *DirectHarm* a dummy variable indicating whether an individual was reportedly injured or maimed during such an attack. *Female* is a fixed effect for being female, and *Village* is a set of village fixed effects. Other pre-treatment covariates are included through $\boldsymbol{X}$, such as: age, whether they were a farmer, herder, merchant or trader, their household size and whether or not they voted in the past. After running

[4] We only describe here the most relevant details, further information is available in Hazlett (2013).

this regression, we find that exposure to violence (*DirectHarm*) is associated with more pro-peace attitudes on *PeaceIndex*.

Despite the claims made regarding "conditionally indiscriminate violence", not all investigators may agree with these arguments, and thus with the assumption of no unobserved confounders. Consider, for example, a fellow researcher who argues that: although bombings were highly indiscriminate and impossible to target finely, perhaps those in the center of the village were more often harmed than those on the periphery. And might not those nearer the center of each village also have different types of attitudes towards peace, on average? This suggests that the author ought to have instead run the model,

$$\text{PeaceIndex} = \hat{\tau}\text{DirectHarm} + \hat{\beta}_f\text{Female} + \text{Village}\hat{\beta}_v + \boldsymbol{X}\hat{\beta} + \hat{\gamma}\text{Center} + \hat{\varepsilon}_{\text{full}} \tag{2}$$

That is, our earlier estimate $\hat{\tau}_{\text{res}}$ would differ from our target quantity $\hat{\tau}$. But how badly? How "strong" a confounder like *Center* would need to be to change our research conclusions? A simple violation of unconfoundedness such as this one can be handled in a relatively straightforward manner by the traditional OVB, as we will see in Section 3.

However, other skeptical researchers may question the within-village indiscriminacy with more elaborate stories, worrying that unobserved factors such as *Wealth* or *Political Attitudes* remain as confounders, perhaps even acting through non-linear functions such as an interaction of these two. Additionally, we may also have expert knowledge that could be used to limit arguments about potential confounding. For example, considering the nature of the attacks and the special role that gender played, one may argue that any within-village confounders are not likely to be as strong as the observed covariate *Female*. We would like to develop tools to answer such questions: how strong would these confounders need to be (acting as a group, possibly with non-linearities) to change our conclusions? And how could we codify and leverage our beliefs about the importance of *Female* to bound the plausible strength of unobserved confounders? In Sections 4 and 5, we show how extending the traditional OVB framework provides answers to these questions.

# 3 Sensitivity in an Omitted Variable Bias Framework

The "omitted variable bias" (OVB) formula is an important part of the mechanics of linear regression models and describes how the inclusion of an omitted covariate changes a coefficient estimate of interest. In this section, we review the traditional OVB approach, and illustrate its use as a simple tool for sensitivity analysis through bivariate contour plots showing how the effect estimate would vary depending upon hypothetical strengths of the confounder. This serves not only as an introduction to the method, but also to highlight limitations of the traditional approach, which we then address in the following sections.

## 3.1 The traditional Omitted Variable Bias

Suppose an investigator wishes to run a linear regression model of an outcome $Y$ on a treatment $D$, controlling for a set of covariates given by $\boldsymbol{X}$ and $Z$, as in

$$Y = \hat{\tau}D + \boldsymbol{X}\hat{\beta} + \hat{\gamma}Z + \hat{\varepsilon}_{\text{full}} \tag{3}$$

where $Y$ is an $(n \times 1)$ vector containing the outcome of interest for each of the $n$ observations and $D$ is an $(n \times 1)$ treatment variable (which may be continuous or binary); $\boldsymbol{X}$ is an $(n \times p)$ matrix

of *observed* (pre-treatment) covariates including the constant; and $Z$ is a single $(n \times 1)$ *unobserved* covariate (we allow a multivariate version of $Z$ in Section 4.5).

However, since $Z$ is unobserved, the investigator is forced instead to estimate a restricted model including $\boldsymbol{X}$ only,

$$Y = \hat{\tau}_{\text{res}} D + \boldsymbol{X}\hat{\beta}_{\text{res}} + \hat{\varepsilon}_{\text{res}} \tag{4}$$

where $\hat{\tau}_{\text{res}}$, $\hat{\beta}_{\text{res}}$ are the estimates of the restricted OLS with only $D$ and $\boldsymbol{X}$, omitting $Z$, and $\hat{\varepsilon}_{\text{res}}$ its corresponding residual.

How does the observed estimate ($\hat{\tau}_{\text{res}}$) compare to the desired estimate, $\hat{\tau}$? Let us define as $\widehat{\text{bias}}$ the difference between these estimates, $\widehat{\text{bias}} := \hat{\tau}_{\text{res}} - \hat{\tau}$, where the hat, $\widehat{(\cdot)}$, clarifies that this quantity is a difference between sample estimates, not the difference between the expectation of a sample estimate and a population value. Using the Frisch-Waugh-Lovell (FWL) theorem (Frisch and Waugh 1933; Lovell 1963 2008) to "partial out" the observed covariates $\boldsymbol{X}$, the classic omitted variable bias solution is

$$\begin{aligned}
\hat{\tau}_{\text{res}} &= \frac{\text{cov}(D^{\perp \boldsymbol{X}},\ Y^{\perp \boldsymbol{X}})}{\text{var}(D^{\perp \boldsymbol{X}})} \\
&= \frac{\text{cov}(D^{\perp \boldsymbol{X}},\ \hat{\tau} D^{\perp \boldsymbol{X}} + \hat{\gamma} Z^{\perp \boldsymbol{X}})}{\text{var}(D^{\perp \boldsymbol{X}})} \\
&= \hat{\tau} + \hat{\gamma}\left(\frac{\text{cov}(D^{\perp \boldsymbol{X}},\ Z^{\perp \boldsymbol{X}})}{\text{var}(D^{\perp \boldsymbol{X}})}\right) \\
&= \hat{\tau} + \hat{\gamma}\hat{\delta}
\end{aligned} \tag{5}$$

where $\text{cov}(\cdot)$ and $\text{var}(\cdot)$ denote the *sample* covariance and variance; $Y^{\perp \boldsymbol{X}}$, $D^{\perp \boldsymbol{X}}$ and $Z^{\perp \boldsymbol{X}}$ are the variables $Y$, $D$ and $Z$ after removing the components linearly explained by $\boldsymbol{X}$ and we define $\hat{\delta} := \frac{\text{cov}(D^{\perp \boldsymbol{X}}, Z^{\perp \boldsymbol{X}})}{\text{var}(D^{\perp \boldsymbol{X}})}$. We then have

$$\widehat{\text{bias}} = \hat{\gamma}\hat{\delta} \tag{6}$$

While elementary, the OVB formula in Equation 6 provides the key intuitions as well as a formulaic basis for a simple sensitivity analysis, letting us assess how the omission of covariates we wished to have controlled for could affect our inferences. Note that it holds *whether or not Equation 3 has a causal meaning*. In applied settings, however, one is typically interested in cases where the investigator has determined that the full regression, controlling for *both* $\boldsymbol{X}$ *and* the unobserved variable $Z$, would have identified the causal effect of $D$ on $Y$; thus, hereafter we will treat $Z$ as an unobserved "confounder" and continue the discussion as if the estimate $\hat{\tau}$, obtained with the inclusion of $Z$, is the desired target quantity.[5]

[5]Conditions that endow regression estimates with causal meaning are extensively discussed in the literature: identification assumptions can be articulated in graphical terms, such as postulating a structural causal model in which $\{\boldsymbol{X}, Z\}$ satisfy the backdoor criterion for identifying the causal effect of $D$ on $Y$ (Pearl 2009); or, equivalently, in counterfactual notation, stating that the treatment assignment $D$ is conditionally ignorable given $\{\boldsymbol{X}, Z\}$, that is $Y_d \perp\!\!\!\perp D|\boldsymbol{X}, Z$, where $Y_d$ denotes the potential outcome of $Y$ when $D$ is *set* to $d$, see Pearl (2009); Angrist and Pischke (2008); Imbens and Rubin (2015). We further note the effects of $D$ on $Y$ may be non-linear, in which case a regression coefficient may be an incomplete summary of the causal effect, see Angrist and Pischke (2008). Finally, indiscriminate inclusion of covariates can induce or amplify bias, see Pearl (2011), Pearl (2012), Middleton et al. (2016), Ding and Miratrix (2015) and Pearl (2015) for related discussions. Here we assume the researcher is interested in the estimates one would obtain from running the regression in Equation 3, controlling for $\boldsymbol{X}$ and $Z$.

### 3.2 Making sense of the traditional OVB

One virtue of the OVB formula is its interpretability. The quantity $\hat{\gamma}$ describes the difference in the linear expectation of the outcome, when comparing individuals that differ by one unit on the confounder, but have the same treatment assignment status as well as the same value for all remaining covariates. That is, in broader terms, $\hat{\gamma}$ describes how looking at different subgroups of the unobserved confounder "impacts" our best linear prediction of the outcome.[6]

By analogy, it would be tempting to think of $\hat{\delta}$ as the estimated marginal "impact" of the confounder on the *treatment*. However, causal interpretation aside, this is incorrect because it refers instead to the reverse regression. That is, $\hat{\delta}$ is the coefficient from the regression $Z = \hat{\delta}D + \boldsymbol{X}\hat{\psi} + \hat{\varepsilon}_Z$, and not the regression of the treatment $D$ on $Z$, and $\boldsymbol{X}$.

That is, it gives the difference in the linear expectation of the confounder, when comparing individuals with the same values for the covariates, but differing by one unit on the treatment. This quantity will be familiar to empirical researchers who have used quasi-experiments in which the treatment is believed to be randomized only conditional on certain covariates $\boldsymbol{X}$. In that case we may then check for "balance" on other (pre-treatment) observables once conditioning is complete. Hence, we can think of $\hat{\delta}$ as the (conditional) imbalance of the confounder with respect to the treatment—or simply "imbalance".

Thus, a useful mnemonic is that the omitted variable bias can be summarized as the unobserved confounder's "impact times its imbalance". It is worth noting that the imbalance component is quite general: whatever the true functional form dictating $\mathbb{E}[Z|D, \boldsymbol{X}]$ (or the treatment assignment mechanism), the only way in which $Z$'s relationship to $D$ enters the bias is captured by its "linear imbalance", parameterized by $\hat{\delta}$. That is, the linear regression of $Z$ on $D$ and $X$ need not reflect the correct expected value of $Z$—rather it serves to capture the aspects of the relationship between $Z$ and $D$ conditionally on $X$ that affects the bias.

### 3.3 Using the traditional OVB for sensitivity analysis

If we know the *signs* of the partial correlations between the confounder with the treatment and the outcome (the same as the signs of $\hat{\gamma}$ and $\hat{\delta}$) we can argue whether our estimate is likely to be underestimating or overestimating the quantity of interest. Arguments using correlational direction is common practice in econometrics work.[7] Often, though, discussing possible direction of the bias is not possible or not sufficient, and magnitude must be considered. How strong would the confounder(s) have to be to change the estimates in such a way to affect the main conclusions of a study?

[6] While a causal interpretation here is tempting, whether this difference in the distribution of the outcome within strata of the confounder can be attributable to a direct causal effect of the former on the latter depends on structural assumptions. If, for example, the "true" outcome model is assumed to be a linear structural equation where strict exogeneity holds, i.e., $Y = \tau D + \boldsymbol{X}\beta + \gamma Z + \varepsilon$ and $\mathbb{E}[\varepsilon|D, \boldsymbol{X}, Z] = 0$, then $\hat{\gamma}$ could be interpreted as an estimate of the direct causal impact of a unit change of the confounder on the expected value of the outcome $Y$, holding the other covariates fixed. In many scenarios, however, this might be unrealistic—since the researcher's goal is to estimate the causal effect of $D$ on $Y$, usually $Z$ is required only to, along with $\boldsymbol{X}$, block the back-door paths from $D$ to $Y$ (Pearl 2009), or equivalently, make the treatment assignment conditionally ignorable. In this case, $\hat{\gamma}$ could reflect not only its causal effect on $Y$, if any, but also other spurious associations not eliminated by standard assumptions. As a heuristic, however, refering to $\hat{\gamma}$ as the marginal "impact" of the confounder on the outcome is useful, as long as the reader keeps in mind the it can only have a direct causal meaning under certain circumstances.

[7] e.g. "Using a similar omitted-variables-type argument, we note that even if there are other confounders that we havent controlled for, those that are positively correlated with private school attendance are likely to be positively correlated with earnings as well. Even if these variables remain omitted, their omission leads the estimates computed with the variables at hand to overestimate the private school premium." (Angrist and Pischke 2017, p.8-9)

### Sensitivity contour plots

A first approach to investigate the sensitivity of our estimate can be summarized by a two-dimensional plot of bias contours parameterized by the two terms $\hat{\gamma}$ and $\hat{\delta}$. Each pair of hypothesized "impact" and "imbalance" parameters corresponds to a certain level of bias (their product), but given an initial treatment effect estimate $\hat{\tau}_{\text{res}}$, we can also relabel the bias levels in terms of the "adjusted" effect estimate, i.e $\hat{\tau} = \hat{\tau}_{\text{res}} - \hat{\gamma}\hat{\delta}$, the estimate from the OLS regression we wish we had run, if we had included a confounder with the hypothesized level of impact and imbalance.

In our running example, a specific confounder we wish we had controlled for is a binary indicator of whether the respondent lived in the center or in the periphery of the village. How strong would this specific confounder have to be in order for its inclusion to substantially affect our conclusions? Figure 1 shows the plot of adjusted estimates for several hypothetical values of impact and imbalance of the confounder *Center*.

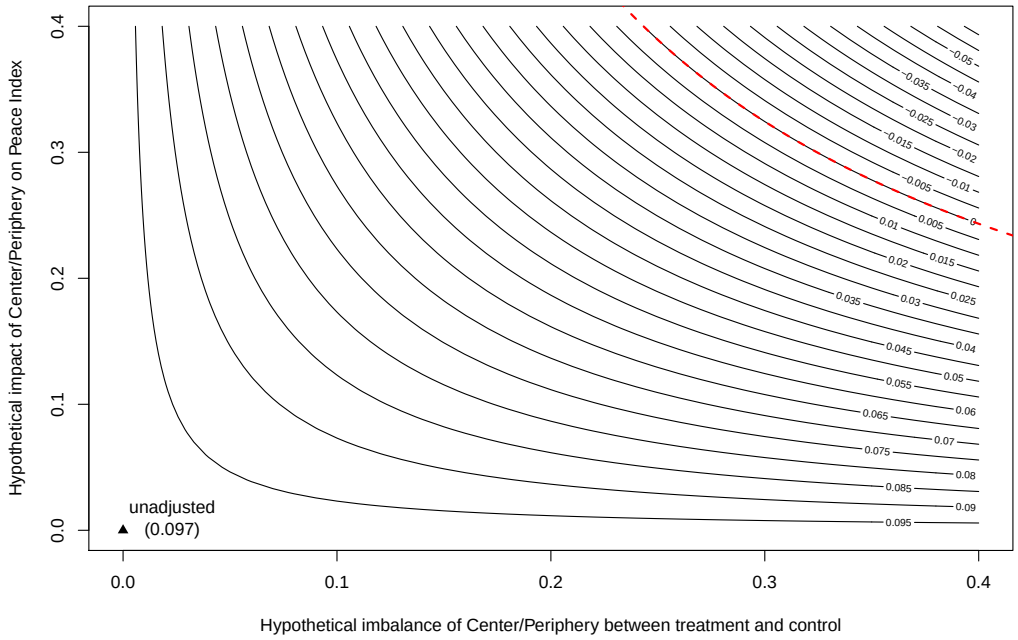


Figure 1: Sensitivity contours of point estimate — traditional OVB

*Note*: Sensitivity analysis with the traditional omitted variable bias formulation. The horizontal axis shows hypothetical values for the imbalance of the confounder *Center* between treated and control units. The vertical axis shows hypothetical values for the impact of the confounder on the outcome. The "heights" shown by the contour lines give the adjusted effect size of the treatment on the outcome—more specifically to our running example, the adjusted effect of *Direct Harm* on *Peace Index*, given hypothetical values for the imbalance and impact of the confounder *Center*. Notice there are four possible quadrants for the sensitivity plot, here we consider only the worst case, in which the bias acts to move effects toward (or even through) zero.

Hypothetical values for the imbalance of the confounder lie on the horizontal axis. For this particular case, they indicate the difference of the average number of individuals living in the center between those who were exposed to direct harm and those who weren't. Values for the hypothetical impact of the confounder on the outcome lie on vertical axis, representing, on average, how attitudes towards peace differ from people living in the center versus those in the periphery of the village, within strata of other covariates.The contour lines of the plot give the adjusted treatment effect at hypothesized values of the impact and imbalance parameters. That is, they show the exact

estimate one would have obtained by running the full regression including a confounder with those hypothetical sensitivity parameters. No other information is required to know how such a confounder would influence the result. Notice here we parameterize the bias in a way that it would always hurt our preferred hypothesis by reducing the effect size.[8]

This plot explicitly reveals the type of prior knowledge one needs to have in order to be able to rule out problematic confounders. As an example, imagine the confounder *Center* has a conditional imbalance as high as 0.25—that is, having controlled for the observed covariates, those who were physically injured were also 25 percentage points more likely to live in the center of the village than those who weren't. With such an imbalance, the plot reveals that the impact of living in the center on the outcome (Peace Index) would have to be over 0.40 in order to bring down the estimated effect of *DirectHarm* to zero.

Determining whether this is good or bad news remains difficult and requires contextual arguments. For instance, one could argue that, given the relatively homogeneous nature of these small villages and that their centers are generally not markedly different in composition than the peripheries, it would be surprising indeed if being in the center was associated with a 0.40 higher expected score on Peace Index (which varies only from 0 to 1). Regardless of whether the investigator can make a clear argument that rules out such confounders, the virtue of sensitivity analysis is that it moves the conversation from one where the investigator seeks to defend "perfect identification" and the critic points out potential confounders, to one where details can be given and discussed about the degree of confounding that would be problematic.

**Shortcomings of the traditional OVB**

The traditional OVB has some benefits: as shown, with sound substantive knowledge about the problem, it's a straightforward exercise. But it also has shortcomings. In the previous example, *Center* was a convenient choice of confounder because it is a binary variable, and the units of measure attached to "impact" and "imbalance" are thus easy to understand as changes in proportions. This is not in general the case. Imagine contemplating confounders such as *Political Attitudes*: in what scale should we measure this? A doubling of that scale would halve the required "impact" and double the required "imbalance". A possible solution is standardizing the coefficients, but this does not help if the goal is to assess the sensitivity of the causal parameter in its original scale.

Furthermore, the traditional OVB, be it standardized or not, does not generalize easily to multiple confounders: how should we assess the effect of confounders *Political Attitudes* and *Wealth*, acting together, perhaps with complex non-linearities? Or, more generally, how should we consider all the other unnamed confounders acting together? Can we benchmark all these confounders against *Female*? Finally, how to obtain the sensitivity of not only the point estimate, but also the standard errors, so that we could examine t-values, p-values or confidence intervals under hypothetical confounders?

# 4 OVB with the partial $R^2$ parameterization

We now consider a reparameterization of the OVB formula in terms of partial $R^2$ values. Our goal is to replace the sensitivity parameters $\hat{\gamma}$ and $\hat{\delta}$ with a pair of parameters that uses an $R^2$ measure

[8] Investigators may also argue that bias would increase the effect size, in the sense that the current estimates are conservative. Our tools apply to these cases as well, the arguments would just work in the opposite direction. For simplicity of exposition, in the paper we will focus on the case where accounting for omitted variable bias reduces the effect size.

to assess the strength of association between the confounder and the treatment and between the confounder and the outcome, both assuming the remaining covariates $\boldsymbol{X}$ have been accounted for. The partial $R^2$ parameterization is scale-free, yet intuitive. It further enables us to construct a number of useful analyses, including: (i) assessing the sensitivity of an estimate to any number or even *all* confounders acting together, possibly non-linearly; (ii) using the same framework to assess the sensitivity of point estimates as well as t-values and confidence intervals; (iii) assessing the sensitivity to extreme-scenarios in which all or a big portion of the unexplained variance of the outcome is due to confounding; (iv) applying contextual information about the research design to bound the strength of the confounders; and (v) presenting these sensitivity results concisely for easy routine reporting, as well as providing intuitive visual tools for finer grained analysis.

### 4.1 Reparameterizing the bias in terms of partial $R^2$

Let $R^2_{Z\sim D}$ denote the (sample) $R^2$ of regressing $Z$ on $D$. Recall that for OLS the following holds, $R^2_{Z\sim D} = \frac{\text{var}(\hat{Z})}{\text{var}(Z)} = 1 - \frac{\text{var}(Z^{\perp D})}{\text{var}(Z)} = \text{cor}(Z, \hat{Z})^2 = \text{cor}(Z, D)^2$, where $\hat{Z}$ are the fitted values given by regressing $Z$ on $D$. Notice the $R^2$ is symmetric, that is, it's invariant to whether one uses the "forward" or the "reverse" regression since $R^2_{Z\sim D} = \text{cor}(Z, D)^2 = \text{cor}(D, Z)^2 = R^2_{D\sim Z}$. Extending this to the case with covariates $\boldsymbol{X}$, we denote the partial $R^2$ from regressing $Z$ on $D$ after controlling for $\boldsymbol{X}$ as $R^2_{Z\sim D|\boldsymbol{X}}$. This has the same useful symmetry, with $R^2_{Z\sim D|\boldsymbol{X}} = 1 - \frac{\text{var}(Z^{\perp \boldsymbol{X},D})}{\text{var}(Z^{\perp X})} = \text{cor}(Z^{\perp \boldsymbol{X}}, D^{\perp \boldsymbol{X}})^2 = \text{cor}(D^{\perp \boldsymbol{X}}, Z^{\perp \boldsymbol{X}})^2 = R^2_{D\sim Z|\boldsymbol{X}}$.

We are now ready to express the bias in terms of partial $R^2$. First, by the FWL theorem,

$$\begin{aligned}
\widehat{\text{bias}} &= \hat{\delta}\hat{\gamma} \\
&= \left(\frac{\text{cov}(D^{\perp \boldsymbol{X}},\ Z^{\perp \boldsymbol{X}})}{\text{var}(D^{\perp \boldsymbol{X}})}\right)\left(\frac{\text{cov}(Y^{\perp \boldsymbol{X},D},\ Z^{\perp \boldsymbol{X},D})}{\text{var}(Z^{\perp \boldsymbol{X},D})}\right) \\
&= \left(\frac{\text{cor}(D^{\perp \boldsymbol{X}},\ Z^{\perp \boldsymbol{X}})\text{sd}(Z^{\perp X})}{\text{sd}(D^{\perp X})}\right)\left(\frac{\text{cor}(Y^{\perp X,D},\ Z^{\perp X,D})\text{sd}(Y^{\perp \boldsymbol{X},D})}{\text{sd}(Z^{\perp \boldsymbol{X},D})}\right) \\
&= \left(\frac{\text{cor}(Y^{\perp \boldsymbol{X},D},\ Z^{\perp \boldsymbol{X},D})\text{cor}(D^{\perp \boldsymbol{X}},\ Z^{\perp \boldsymbol{X}})}{\frac{\text{sd}(Z^{\perp \boldsymbol{X},D})}{\text{sd}(Z^{\perp \boldsymbol{X}})}}\right)\left(\frac{\text{sd}(Y^{\perp \boldsymbol{X},D})}{\text{sd}(D^{\perp \boldsymbol{X}})}\right) \qquad (7)
\end{aligned}$$

Noting that $\text{cor}(Y^{\perp \boldsymbol{X},D}, Z^{\perp \boldsymbol{X},D})^2 = R^2_{Y\sim Z|\boldsymbol{X},D}$, that $\text{cor}(Z^{\perp \boldsymbol{X}},\ D^{\perp \boldsymbol{X}})^2 = R^2_{D\sim Z|\boldsymbol{X}}$, and that $\frac{\text{var}(Z^{\perp \boldsymbol{X},D})}{\text{var}(Z^{\perp \boldsymbol{X}})} = 1 - R^2_{Z\sim D|\boldsymbol{X}} = 1 - R^2_{D\sim Z|\boldsymbol{X}}$, we can write 7 as:

$$|\widehat{\text{bias}}| = \sqrt{\frac{R^2_{Y\sim Z|\boldsymbol{X},D}\ R^2_{D\sim Z|\boldsymbol{X}}}{1 - R^2_{D\sim Z|\boldsymbol{X}}}}\left(\frac{\text{sd}(Y^{\perp \boldsymbol{X},D})}{\text{sd}(D^{\perp \boldsymbol{X}})}\right) \qquad (8)$$

Equation 8 rewrites the OVB formula in terms that more conveniently rely on partial $R^2$ measures of association rather than raw regression coefficients. Investigators may be interested in how confounders alter inference as well, so we also examine the standard error. Let df denote the regression's degrees of freedom (for the restricted regression actually run). Noting that

$$\text{se}(\hat{\tau}_{\text{res}}) = \frac{\text{sd}(Y^{\perp \boldsymbol{X},D})}{\text{sd}(D^{\perp \boldsymbol{X}})}\sqrt{\frac{1}{\text{df}}} \tag{9}$$

$$\text{se}(\hat{\tau}) = \frac{\text{sd}(Y^{\perp \boldsymbol{X},D,Z})}{\text{sd}(D^{\perp \boldsymbol{X},Z})}\sqrt{\frac{1}{\text{df}-1}} \tag{10}$$

whose ratio is

$$\frac{\text{se}(\hat{\tau})}{\text{se}(\hat{\tau}_{\text{res}})} = \left(\frac{\text{sd}(Y^{\perp \boldsymbol{X},D,Z})}{\text{sd}(Y^{\perp \boldsymbol{X},D})}\right)\left(\frac{\text{sd}(D^{\perp \boldsymbol{X}})}{\text{sd}(D^{\perp \boldsymbol{X},Z})}\right)\sqrt{\frac{\text{df}}{\text{df}-1}} \tag{11}$$

We obtain the expression for the standard error of $\hat{\tau}$

$$\text{se}(\hat{\tau}) = \text{se}(\hat{\tau}_{\text{res}})\sqrt{\frac{1-R^2_{Y\sim Z|\boldsymbol{X},D}}{1-R^2_{D\sim Z|\boldsymbol{X}}}\left(\frac{\text{df}}{\text{df}-1}\right)} \tag{12}$$

Moreover, with this we can further see the bias as:

$$\widehat{|\text{bias}|} = \text{se}(\hat{\tau}_{\text{res}})\sqrt{\frac{R^2_{Y\sim Z|\boldsymbol{X},D}\ R^2_{D\sim Z|\boldsymbol{X}}}{1-R^2_{D\sim Z|\boldsymbol{X}}}(\text{df})} \tag{13}$$

## 4.2 Making sense of the partial $R^2$ parameterization

Equations 12 and 13 form the basis of the sensitivity exercises regarding both the point estimate and the standard error, with sensitivity parameters in terms of $R^2_{Y\sim Z|\boldsymbol{X},D}$ and $R^2_{D\sim Z|\boldsymbol{X}}$. These formulae are computationally convenient—the only data dependent parts are the standard error of $\hat{\tau}_{\text{res}}$ and the regression's degrees of freedom, which are already reported by most regression software. In this section, we provide remarks that help making sense of these results, revealing their simplicity in terms of regression anatomy: the (relative) bias of the point estimate is determined by a ratio of two partial Cohen's $f$ statistics, and the change in variance is given by a product of three interpretable components.[9]

### Sensitivity of the point estimate

In the partial $R^2$ parameterization, the relative bias, $\left|\frac{\widehat{\text{bias}}}{\hat{\tau}_{\text{res}}}\right|$, has a simple form:

$$\text{relative bias} = \frac{\overbrace{|R_{Y\sim Z|\boldsymbol{X},D} \times f_{D\sim Z|\boldsymbol{X}}|}^{\text{bias factor}}}{\underbrace{|f_{Y\sim D|\boldsymbol{X}}|}_{\text{partial f of D with Y}}} = \frac{BF}{|f_{Y\sim D|\boldsymbol{X}}|} \tag{14}$$

The numerator of the relative bias contains the partial Cohen's $f$ of the confounder with the treatment, "adjusted" by the partial correlation of that confounder with the outcome.[10] Collectively

[9] See appendix A.1 for details.

[10] Cohen's $f^2$ can be written as $f^2 = R^2/(1-R^2)$, so, for example, $f^2_{D\sim Z|\boldsymbol{X}} = R^2_{D\sim Z|\boldsymbol{X}}/(1-R^2_{D\sim Z|\boldsymbol{X}})$.

this numerator could be called the "bias factor" of the confounder, $BF = |R_{Y\sim Z|\boldsymbol{X},D} \times f_{D\sim Z|\boldsymbol{X}}|$, which is determined entirely by the two sensitivity parameters ($R^2_{Y\sim Z|\boldsymbol{X},D}$, $R^2_{D\sim Z|\boldsymbol{X}}$). To determine the size of the relative bias, this is compared to how much variation of the outcome is uniquely explained by the treatment assignment, in the form of the partial $f$ of the treatment with the outcome. Computationally, the partial $f$ equals the estimate's t-value divided by $\sqrt{\text{df}}$. This allows one to easily assess sensitivity to any confounder with a given pair of partial $R^2$ values, see Table 2 in Appendix A.4 for an illustrating procedure.

Equation 14 also reveals that, given a particular confounder (which will fix $BF$), the only property needed to determine the robustness of a regression estimate against that confounder is the partial $R^2$ of the treatment with the outcome (via $f_{Y\sim D|\boldsymbol{X}}$). This serves to reinforce the fact that robustness to confounding is an identification problem, impervious to sample size considerations. While t-values and p-values might be informative with respect to the statistical uncertainty (in a correctly specified model), robustness to mispecification is determined by the share of variation of the outcome the treatment uniquely explains.

A subtle but useful property of the partial $R^2$ parameterization is that it reveals an asymmetry in the role of the components of the bias factor. In the traditional OVB formulation, the bias is simply a product of two terms with the same importance. The new formulation breaks this symmetry: the effect of the partial $R^2$ of the confounder with the outcome on the bias factor is bounded at one. By contrast, the effect of partial $R^2$ of the confounder with the treatment on the bias factor is unbounded (via $f_{D\sim Z|\boldsymbol{X}}$). This allows us to consider extreme scenarios, in which we suppose the confounder explains *all* of the left-out variation of the outcome, and see what happens as we vary the partial $R^2$ of the confounder with the treatement (Section 5.3).

**Sensitivity of the variance**

How the confounder affects the variance has a straightforward interpretation as well. The relative change in the variance, $\frac{\text{var}(\hat{\tau})}{\text{var}(\hat{\tau}_{\text{res}})}$, can be decomposed into three components,

$$\begin{aligned}\text{relative change in variance} &= \overbrace{\left(1 - R^2_{Y\sim Z|\boldsymbol{X},D}\right)}^{\text{VRF}} \underbrace{\left(\frac{1}{1 - R^2_{D\sim Z|\boldsymbol{X}}}\right)}_{\text{VIF}} \overbrace{\left(\frac{\text{df}}{\text{df} - 1}\right)}^{\text{change in df}} \\ &= \text{VRF} \times \text{VIF} \times \text{change in df} \end{aligned} \tag{15}$$

That is, including the confounder in the regression reduces the variance of the coefficient of $D$ by reducing the residual variance of $Y$ (variance reduction factor—VRF). On the other hand, it raises the variance of the coefficient via its partial correlation with the treatment (the traditional variance inflation factor—VIF). Finally, the degrees of freedom must be adjusted. The overall relative change of the variance is simply the product of these three components.

## 4.3 Sensitivity statistics for routine reporting

Detailed sensitivity analyses can be conducted using the previous results, as we will show in the next section. However, widespread use of sensitivity analyses may also be facilitated by providing simple sensitivity measures that quickly describes the overall sensitivity of an estimate to unobserved confounding. These two measures have two uses: (i) they can be routinely reported in standard regression tables, making the discussion of sensitivity to unobserved confounding more accessible and

standard practice; and, (ii) they can easily be computed from standard quantities found on a regression table, allowing readers and reviewers to initiate the discussion about unobserved confounders when reading papers that did not formally assess sensitivity.

**The robustness value**

The first quantity we propose is the *robustness value* (RV), which conveniently summarizes the types of confounders that would problematically change our research conclusions. Consider a confounder with equal association to the treatment and the outcome, i.e. $R^2_{Y\sim Z|\boldsymbol{X},D} = R^2_{D\sim Z|\boldsymbol{X}} = RV_q$. The $RV_q$ describes how strong that association must be in order to reduce the estimated effect by $(100 \times q)\%$. By Equation 14 (see also Appendix A.1),

$$RV_q = \frac{1}{2}\left(\sqrt{f_q^4 + 4f_q^2} - f_q^2\right) \tag{16}$$

where $f_q := q|f_{Y\sim D|\boldsymbol{X}}|$ is the partial Cohen's $f$ of the treatment with the outcome multiplied by the proportion of reduction $q$ on the treatment coefficient which would be deemed problematic. Confounders that explain $RV_q\%$ both of the treatment and the outcome are sufficiently strong to change the point estimate in problematic ways, while confounders with neither association greater than $RV_q\%$ are not.

The $RV$ thus offers an interpretable sensitivity measure that summarizes how robust the point estimate is to unobserved confounding. A robustness value close to one means the treatment effect can handle strong confounders explaining almost all variation of the treatment and the outcome. On the other hand, a robustness value close to zero means that even very weak confounders could eliminate the results. Note that the $RV$ can be easily computed in any regression table, recalling that $f_{Y\sim D|\boldsymbol{X}}$ can be obtained by simply dividing the treatment coefficient t-value by the square-root of the degrees of freedom.

With minor adjustment, robustness values can also be obtained for t-values, or lower and upper bounds of confidence intervals as well. Let $|t^{\alpha}_{\mathrm{df}-1}|$ denote the t-value threshold for a t-test with significance level of $\alpha$ and $\mathrm{df}-1$ degrees of freedom. Now construct an adjusted $f_{q,\alpha}$, accounting for both the proportion of reduction $(q)$ of the point estimate and the boundary below which statistical significance is lost at the level of $\alpha$ ,

$$f_{q,\alpha} := q|f_{Y\sim D|\boldsymbol{X}}| - \frac{|t^{\alpha}_{\mathrm{df}-1}|}{\sqrt{\mathrm{df}-1}} \tag{17}$$

Then, a confounder with a partial $R^2$ of,

$$RV_{q,\alpha} = \frac{1}{2}\left(\sqrt{f_{q,\alpha}^4 + 4f_{q,\alpha}^2} - f_{q,\alpha}^2\right) \tag{18}$$

both with the treatment and with the outcome is sufficiently strong to make the adjusted t-test not reject the hypothesis $H_0 : \tau = (1-q)|\hat{\tau}_{\mathrm{res}}|$ at the $\alpha$ level (or, equivalently, to make the adjusted $1-\alpha$ confidence interval include $(1-q)|\hat{\tau}_{\mathrm{res}}|$). Note that, since we are considering sample uncertainty, $RV_{q,\alpha}$ is a more conservative measure than $RV_q$. If one picks $|t^{\alpha}_{\mathrm{df}-1}| = 0$ then $RV_{q,\alpha}$ reduces to $RV_q$. Also, for fixed $|t^{\alpha}_{\mathrm{df}-1}|$, $RV_{q,\alpha}$ converges to $RV_q$ when the sample size grows to infinity. See Appendix A.1 for details.[11]

[11] For convenience, we refer to the $RV_q$ or $RV_{q,\alpha}$ with $q = 1$ as simply the $RV$ or $RV_\alpha$

**The $R^2_{Y\sim D|\boldsymbol{X}}$ as an extreme scenario analysis**

The second measure we propose is the proportion of variation in the outcome uniquely explained by the treatment—$R^2_{Y\sim D|\boldsymbol{X}}$. Let us consider an extreme confounder that explains *all* residual variance of the outcome, i.e., $R_{Y\sim Z|\boldsymbol{X},D} = 1$. By Equation 14, to bring down the estimated effect to zero (relative bias = 1), we would have

$$|f_{D\sim Z|\boldsymbol{X}}| = |f_{Y\sim D|\boldsymbol{X}}| \implies R^2_{D\sim Z|\boldsymbol{X}} = R^2_{Y\sim D|\boldsymbol{X}} \tag{19}$$

Thus, $R^2_{Y\sim D|\boldsymbol{X}}$ is not only the determinant of the robustness of the treatment effect coefficient, but it is also a measure of its robustness to an extreme sensitivity analysis scenario. Specifically, in the extreme scenario where the confounder explains all the residual variance of the outcome, to bring down the estimated effect to zero, the partial $R^2$ of the confounder with the treatment would need to exactly equal the partial $R^2$ of the treatment with the outcome.

## 4.4 Bounding the strength of the confounder using observed covariates

Arguably, the most difficult part of a sensitivity analysis is taking the description of a confounder that would be problematic from the formal analysis and reasoning about whether such a confounder might exist in one's study given its design and the investigator's contextual knowledge. This has led some authors to propose informal benchmarking procedures, using statistics of the observed covariates to help researchers "calibrate" their intuitions about the strength of the unobserved confounder (Imbens 2003; Hosman et al. 2010; Blackwell 2013; Dorie et al. 2016; Middleton et al. 2016).

However, this informal benchmarking practice can lead to widely incorrect conclusions, even in the ideal case where researchers do have the correct knowledge about how $Z$ compares to $\boldsymbol{X}$. This happens because the estimates of how the observed covariates are related to the outcome are themselves affected by the omission of $Z$, regardless of whether one assumes $Z$ to be independent of $\boldsymbol{X}$ (see Appendix A.2). We thus advise against informal benchmarking procedures, and previous studies relying upon these methods may warrant revisiting.

Here we introduce a novel approach that allows researchers to use their expert knowledge about observed covariates to *formally bound* the strength of the unobserved confounder under clear assumptions. The method relies on contextual knowlege about the relative power of observed covariates vis-a-vis unobserved covariates to explain the observed variation of the treatment assignment and the outcome. We offer three main alternatives to bound the strength of the unobserved confounder, by judging: (i) how the *total* $R^2$ of the confounder compares with the *total* $R^2$ of a group of observed covariates; (ii) how the *partial* $R^2$ of the confounder compares with the *partial* $R^2$ of a group of observed covariates, having taken into account the explanatory power of remaining observed covariates; or, (iii) how the *partial* $R^2$ of the confounder compares with the *partial* $R^2$ of a group of observed covariates, having taken into account the explanatory power of remaining observed covariates *and* the treatment assignment.

The choice of the benchmarking procedure depends on which of these the researcher prefers to use and can most soundly reason about in their own research. In our running example, for instance, we can think about violence within the village and how it may be distributed. Thus, we make arguments relative to the strength of the covariate *Female* in explaining the treatment and the outcome, *after* removing the variation due to *Village* (and the other covariates). With this in mind, in the main text we illustrate the third type of benchmark, and readers can refer to Appendix A.3 for discussion

of the other two variants.[12]

Assume $Z \perp \boldsymbol{X}$, or, equivalently, consider only the part of $Z$ not linearly explained by $\boldsymbol{X}$. Now suppose the researcher believes she has measured the key determinants of the outcome and treatment assignment process, in the sense that the omitted variables cannot explain as much variance (or cannot explain a large multiple of the variance) of $D$ or $Y$ in comparison to the variance explained by particular observed covariate $X_j$. More formally, define $k_D$ and $k_Y$ as,

$$k_D := \frac{R^2_{D\sim Z|\boldsymbol{X}_{-j}}}{R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim Z|\boldsymbol{X}_{-j},D}}{R^2_{Y\sim X_j|\boldsymbol{X}_{-j},D}} \tag{20}$$

Where $\boldsymbol{X}_{-j}$ represents the vector of covariates $\boldsymbol{X}$ excluding $X_j$. That is, $k_D$ indexes how much variance of the treatment assignment the confounder explains relative to how much $X_j$ explains (after controlling for the remaining covariates). To make things concrete, for instance, if the researcher believes the omission of $X_j$ would result in a larger mean squared error of the treatment assignment regression than the omission of $Z$, this equals the claim $k_D \leq 1$. The same reasoning applies to $k_Y$.

Given parameters $k_D$ and $k_Y$, we can rewrite the strength of the confounders as,

$$R^2_{D\sim Z|\boldsymbol{X}} = k_D f^2_{D\sim X_j|\boldsymbol{X}-j}, \qquad R^2_{Y\sim Z|\boldsymbol{X},D} \leq \eta^2 f^2_{Y\sim X_j|\boldsymbol{X}_{-j},D} \tag{21}$$

where $\eta$ is a scalar which depends on both $k_Y$ and $k_D$ (see Appendix A.3 for details.). These equations allow us to investigate the maximum effect a confounder at most "k times" as strong as a particular covariate $X_j$ would have on the coefficient estimate. These results are also tight, in the sense that we can always find a confounder that makes the second inequality an equality. Further, certain values for $k_D$ and $k_Y$ may be ruled out by the data (for instance, if $R^2_{D\sim X_j|\boldsymbol{X}_{-j}} = 50\%$ then $k_D$ must be less than 1).

Our bounding exercises can be extended to any subset of the covariates. For instance, the researcher can bound the effect of a confounder as strong as *all* covariates $\boldsymbol{X}$. The method is also flexible to allow different subgroups of covariates to bound $R^2_{D\sim Z|\boldsymbol{X}}$ and $R^2_{Y\sim Z|\boldsymbol{X},D}$— thus, if a group of covariates $\boldsymbol{X}_1$ is known to be the most important driver of selection to treatment, and another group of covariates $\boldsymbol{X}_2$ is known to be the most important determinant of the outcome, the researcher can easily exploit this fact. For readers familiar with Oster (2017), we discuss the differences between these in Section 6.1.

### 4.5 Sensitivity to multiple confounders

The previous results let us assess the bias caused by a single confounder. Fortunately, they can also be used to provide *upper bounds* in the case of *multiple* unobserved confounders, and the argument is relatively straightforward.[13]. Allowing $\boldsymbol{Z}$ to be a set (matrix) of confounders (and $\gamma$ its coefficient vector), the ful equation we wished we had estimated becomes:

$$Y = \hat{\tau}D + \boldsymbol{X}\hat{\beta} + \boldsymbol{Z}\hat{\gamma} + \hat{\varepsilon}_{\text{full}} \tag{22}$$

[12]Our software currently also implements this type of benchmark, though options to employ the other two are forthcoming.

[13]See Hosman et al. (2010), Section 4.1, for an alternative proof.

Now consider the single variable $Z^* = \boldsymbol{Z}\hat{\gamma}$. The bias caused by omitting $\boldsymbol{Z}$ is the same as omitting the linear combination $Z^*$, and one can think about the effect of multiple confounders in terms of this single confounder. Estimating the regression with $\boldsymbol{X}$ and $Z^*$ instead of $\boldsymbol{X}$ and $\boldsymbol{Z}$ gives the same results for $\hat{\tau}$:

$$Y = \hat{\tau}D + \boldsymbol{X}\hat{\beta} + Z^* + \hat{\varepsilon}_{\text{full}} \tag{23}$$

Accordingly, $Z^*$ has the same partial $R^2$ with the outcome as the full set $\boldsymbol{Z}$. However, the partial $R^2$ of $Z^*$ with the treatment must be less than or equal to the partial $R^2$ of $\boldsymbol{Z}$ with the treatment—this follows simply because the choice of the linear combination $\hat{\gamma}$ is the one the maximizes the $R^2$ with the outcome, and not with the treatment. Hence, the bias caused by a multivariate $\boldsymbol{Z}$ must be less than or equal the bias computed using Equation 13.

A similar reasoning can be applied to the standard errors. Since the effective partial $R^2$ of the linear combination $Z^*$ with the treatment is less than that of $\boldsymbol{Z}$, simply modifying sensitivity Equation 12 to account for the correct degrees of freedom ($\text{df} - k$ instead of $\text{df} - 1$) will give conservative adjusted standard errors for a multivariate confounder. From a practical point of view, however, we note that further correction of the degrees of freedom might be an unnecessary formality—we are performing a hypothetical exercise, and one can always imagine to have measured $Z^*$.

Finally, note the set of confounders $\boldsymbol{Z}$ is arbitrary, thus it accommodates nonlinear confounders as well as misspecification of the functional form of the observed covariates $\boldsymbol{X}$. To illustrate the point, let $Y = \tau D + \beta_1 X + \gamma_1 Z + \gamma_2 Z^2 + \gamma_3 (Z \times X) + \gamma_4 X^2 + \varepsilon$, and imagine the researcher did not measure $Z$ and did not consider that $X$ could also enter the equation with a squared term. Now just call $\boldsymbol{Z} = (Z_1 = Z, Z_2 = Z^2, Z_3 = Z \times X, Z_4 = X^2)$ and all the previous arguments follow.

# 5 Using the partial $R^2$ parameterization for sensitivity analysis

Returning to our running example of violence in Darfur, we illustrate how these tools can be deployed in an effort to answer the following questions: (i) How strong would a particular confounder (or group of confounders) have to be to change our conclusions? (ii) In a worst case scenario, how vulnerable is our result to *many* or *all* unobserved confounders acting together, possibly non-linearly? (iii) Are the confounders that would alter our conclusions plausible, or at least how strong would they have to be relative to observed covariates?

## 5.1 Proposed minimal reporting: Robustness Value, $R^2_{Y\sim D|\mathbf{X}}$ and Bounds

Table 1 illustrates the type of reporting we propose should accompany linear regression models used for causal inference with observational data. Along with traditionally reported statistics, we propose researchers present (i) the partial $R^2$ of the treatment with the outcome, and (ii) the robustness value, $RV$, both for where the point estimate and the confidence interval would cross zero (or another meaningful reference value). Finally, in order to help judgment, we additionally encourage researchers to provide plausible bounds on the strength of the confounder, either based upon comparison of *meaningful* covariates suggested by the research context and design, as discussed in Section 4.4, or by relying on theory and previous literature where possible.

For our running example of violence in Darfur, Table 1 shows an augmented regression table, including the robustness value ($RV$) of the *Directly Harmed* coefficient, 13.9%. This means that unobserved confounders explaining at least 13.9% of the residual variance of both the treatment and the outcome would explain away the estimated treatment effect. It also means that any confounder

Outcome: *Peace Index*

| Treatment: | Est. | SE | t-value | $R^2_{Y\sim D\|\boldsymbol{X}}$ | $RV$ | $RV_{\alpha=0.05}$ |
|---|---|---|---|---|---|---|
| *Directly Harmed* | 0.097 | 0.023 | 4.18 | 2.2% | 13.9% | 7.6% |
| df = 783, Bound ($Z$ as strong as *Female*): $R^2_{Y\sim Z\|\boldsymbol{X},D}$ = 12%, $R^2_{D\sim Z\|\boldsymbol{X}}$ = 1% | | | | | | |

Table 1: Proposed minimal reporting on sensitivity to unobserved confounders

explaining less than 13.9% of the residual variance of both of the treatment and of the outcome would not be strong enough to bring down the estimated effect to zero. For cases where one association is over 13.9% and the other is below, we will need to conduct additional analyses (illustrated below) to make a conclusion. Nevertheless, the $RV$ provides a quick, meaningful reference point for understanding overall sensitivity.

Similarly, adjusting for confounding may not bring the estimate to zero, but rather into a range where it is no longer "statistically significant". The robustness value accounting for statistical significance is also shown in the table, $RV_{\alpha=0.05}$. For a significance level of 5%, the robustness value goes down from 13.9% to 7.6%—that is, confounders would need to be only about half as strong to make the estimate not "statistically significant". Next, the partial $R^2$ of the treatment with the outcome, $R^2_{Y\sim D|\boldsymbol{X}}$, in Table 1 gives a sensitivity analysis for an extreme scenario: if confounders explained 100% of the residual variance of the outcome, they would need to explain at least 2.2% of the residual variance of the treatment to bring down the estimated effect to zero.

Ruling out confounders of the strengths revealed to be problematic requires, in every case, substantive knowledge of the research context and design. If one can argue to have measured the most important covariates, it is possible to bound the strength of the confounder (Section 4.4) and judge where it falls relative to these quantities. The right corner of Table 1 shows the strength of association that a confounder as strong as *Female* would have: $R^2_{Y\sim Z|\boldsymbol{X},D} = 12\%$ and $R^2_{D\sim Z|\boldsymbol{X}} = 1\%$. As the robustness value is higher than either quantity, the table readily reveals that such a confounder could not fully eliminate the point estimate. In addition, since the bound for $R^2_{D\sim Z|\boldsymbol{X}}$ is less than $R^2_{Y\sim D|\boldsymbol{X}} = 2.2\%$, a "worst case confounder" explaining all of the left-out variance of the outcome and as strongly associated with the treatment as *Female* would not eliminate the estimated effect either.

Expert knowledge and a sound research design are required to make such comparisons meaningful: in our running example, a reasonable argument can be made that gender is one of the most visually apparent characteristics of an individual during the attacks, and that if violence was targeted, this was likely the most important basis for it—in particular, many women were sexually assalted during *Janjaweed* attacks. If one can argue that total confounding as strong as *Female* is implausible, the results show it cannot completely account for the observed estimated effect.

These sensitivity exercises are exact when considering a single linear unobserved confounder and are conservative for multiple unobserved confounders, possibly acting non-linearly—this includes the explanatory power of *all left out factors*, even misspecification of the functional form of observed covariates. It is worth pointing out that sensitivity to any arbitrary confounder with a given pair of partial $R^2$ can also be easily computed with the information on the table, see example in Appendix A.4.

### 5.2 Sensitivity contour plots with partial $R^2$: estimates and t-values

The next step is to refine the analysis with tools that visually demonstrate how confounders of different types would affect point estimates and $t$-values,[14] while showing where bounds on such confounders would fall under different assumptions on how unobserved confounders compare to observables.

Perhaps the first plot investigators would examine would be one similar to Figure 1, but now in the partial $R^2$ parameterization (Figure 2). The horizontal axis describes the fraction of the residual variation in the treatment (partial $R^2$) explained by the confounder; the vertical axis describes the fraction of the residual variation in the outcome explained by the confounder. The contours show the adjusted estimate that would be obtained for an unobserved confounder (in the full model) with the hypothesized values of the sensitivity parameters (assuming the direction of the effects hurts our preferred hypothesis). While in the contour plot of the traditional OVB we focused on a specific binary confounder—*Center*—the contour plot with the partial $R^2$ parameterization allows us to assess sensitivity to any confounder, irrespective of its unit of measure. Additionally, since the sensitivity equations give an upper bound for the multivariate case, the same plot can be used to assess the sensitivity to any *group* of confounders, here including non-linear terms, such as the example of *Political Attitudes* and *Wealth* acting together. Notice that the RV is simply the point where, if we picked a contour of interest (such as where the effect equals zero), we determine where this crosses the 45-degree line, as a means of summarizing the sensitivity parameters that correspond to that point.

Further, the bounding exercise results in points on the plot showing the bounds on the partial $R^2$ of the unobserved confounder if it were $k$ times "as strong" as the observed covariate *Female*. The first point shows the bounds for a confounder (or group of confounders) as strong as *Female*, as was also shown in Table 1. A second reference point shows the bounds for confounders *twice* as strong as *Female*, and finally the last point bounds the strength of confounders *three times* as strong as *Female*. The plot reveals that the *sign of the point estimate* is still relatively robust to confounding with such strengths, although the magnitude would be reduced to 77%, 55% and 32% of the original estimate, respectively.

Moving to inferential concerns, Figure 3 now shows the sensitivity of the *t-value* of the treatment effect. As expected, the sensitivity of the treatment effect changes when considering the t-value — as we move along the horizontal axis, not only the adjusted effect reduces, but we also get larger standard-errors due to the variance inflation factor of the confounder. If we take the t-value of 2 as our reference (the usual approximate value for a 95% confidence interval), the plot reveals the statistical significance of *Directly Harmed* is robust to a confounder as strong as, or twice strong as *Female*. However, whereas confounders *three times* as strong as *Female* would not erode the point estimate to zero, we cannot guarantee the estimate would remain "statistically significant" at the 5% level.

Altogether, these bounding exercises lead to the questions: are such confounders plausible? Do we think it possible that a confounder might exist that is three times as strong as *Female* be? If so, what are they? While one may not have complete confidence in answering such questions, we have moved the discussion from a qualitative argument about whether any confounding is possible to a more disciplined, quantitative argument that entices researchers to think about possible threats to their design.

[14] Here we show only the plots for point estimates and t-values. P-values can be obtained directly from the t-values and confidence interval end-points by adjusting the estimate with the appropriate multiple of the standard-errors.

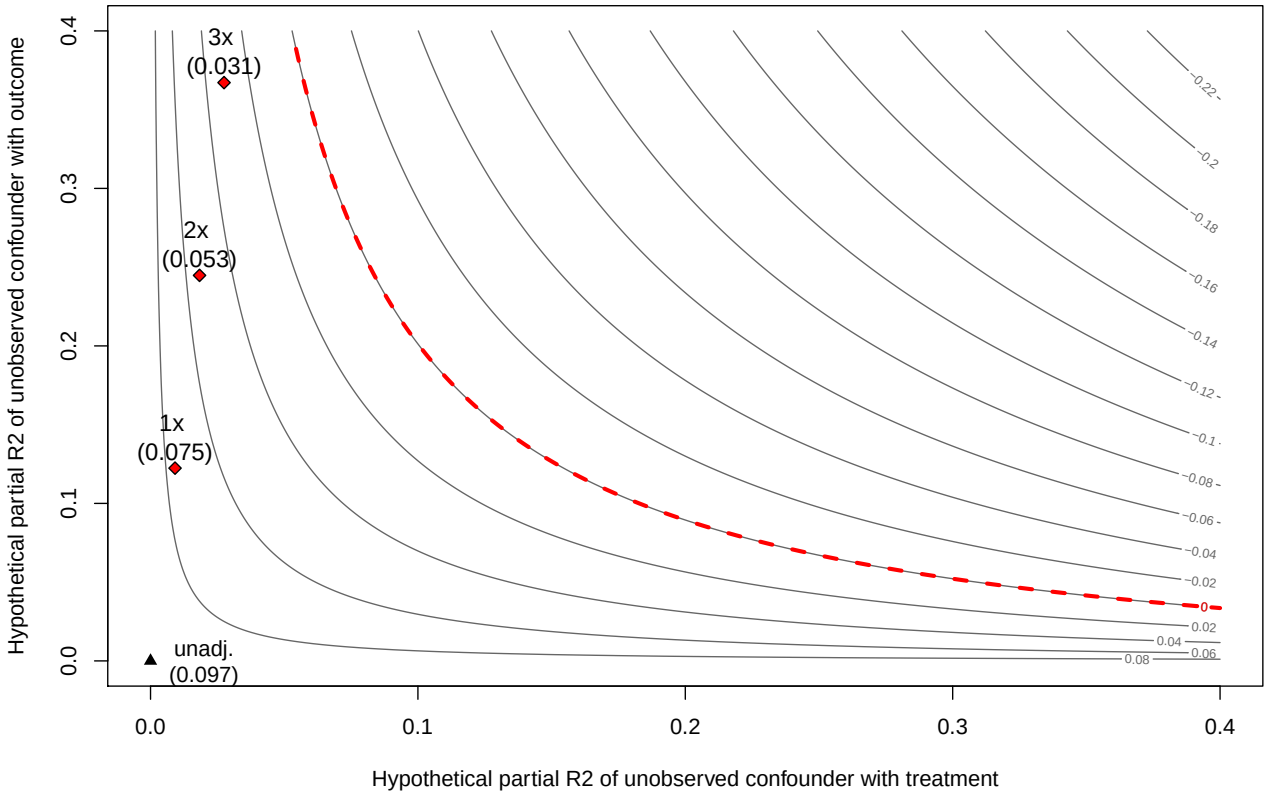


Figure 2: Sensitivity contours of point estimate with benchmark bounds (partial $R^2$)

**Note**: Sensitivity analysis using partial $R^2$ including benchmark bounds using observed covariates. The horizontal axis shows hypothetical values of the partial $R^2$ of the unobserved confounder(s) with the treatment. This can be interpreted as the percentage of the residual variance of the treatment explained by the confounder. The vertical axis shows hypothetical values of the partial $R^2$ of the unobserved confounder(s) with the outcome. Again, this can be interpreted as the percentage of the residual variance of the outcome explained by the confounder. The contour levels represent the adjusted estimates of the treatment effect. The reference points (in red) are bounds on the partial $R^2$ of the unobserved confounder if it were $k$ times "as strong" as the observed covariate *Female*. They show what would be the maximum bias caused by orthogonal unobserved confounder(s) if it had the same or less predictive power than *Female*, both with the treatment and with the outcome (see Appendix A.3 for details). In this particular example, the point estimate of the treatment effect would be robust to a confounder one time, twice or three times as strong as *Female*. For the case of multiple, non-linear, unobserved confounders, these adjusted estimates are conservative—that is, the values are the worst bias multiple, non-linear, confounders could cause for a given pair of partial $R^2$.

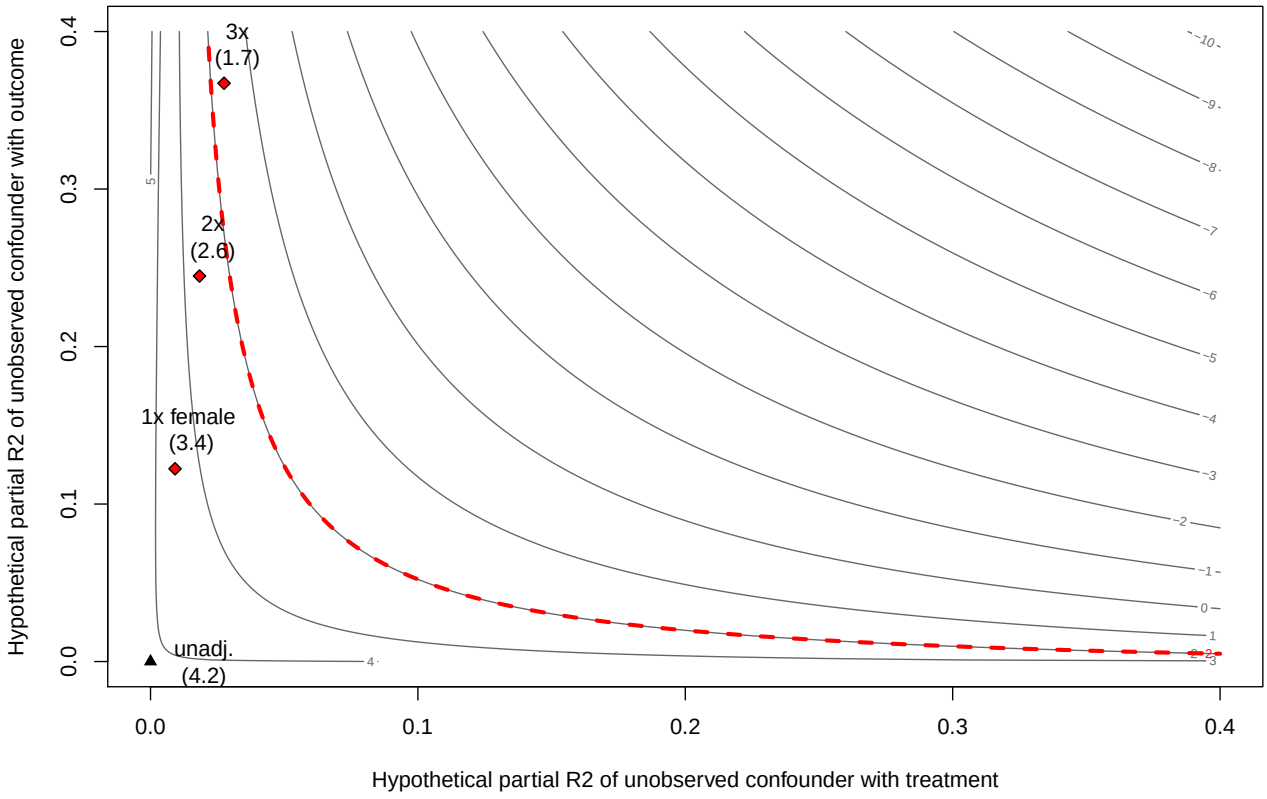


Figure 3: Sensitivity contours of t-value with benchmark bounds (partial $R^2$)

**Note**: Sensitivity analysis of the t-value using partial $R^2$ and including benchmark bounds using observed covariates. Axes are defined as before (see Figure 2), but contour levels now show the adjusted *t statistics* of the treatment effect for given pairs of partial $R^2$ of the confounder. The reference points show the same bounds on the strength of the unobserved confounder if it were $k$ times "as strong" as the observed covariate *Female*. Notice that the t-value is robust to a confounder as strong or twice as strong as *Female*. However, whereas the point estimate was robust to a confounder three times as strong as female, in this example we cannot rule out that such a confounder would reduce the t-value below the usual 5% significance level threshold.

### 5.3 Sensitivity plots of extreme scenarios

Investigators may not always be equipped to argue that confounding is very limited in its association with the outcome. In such cases, exploring sensitivity analysis to extreme-scenarios is still an option. If we set $R^2_{Y\sim Z|\boldsymbol{X},D}$ to one or some other conservative value, how strongly associated with with the treatment the confounder would have to be, in order to problematically change our estimate?

Applying this to our running example, results are shown in Figure 4. The solid curve represents the case where unobserved confounder(s) *explain all the left-out residual variance of the outcome*. On the vertical axis we have the adjusted treatment effect, starting from the case with no bias and going down as the bias increases, reducing the estimate; the horizontal axis shows the partial $R^2$ of the confounder with the treatment. In this *extreme scenario*, as we have seen, $R^2_{D\sim Z|\boldsymbol{X}}$ would need to be exactly the same as the partial $R^2$ of the treatment with the outcome to bring down the estimated effect to zero—that is, it would need to be at least 2.2%, a value below the bound for a confounder twice as strong as *Female*, which in this case is arguably one of the strongest predictors of the treatment assignment. In most circumstances, considering the worst case scenario of $R^2_{Y\sim Z|\boldsymbol{X},D} = 1$ might be needlessly conservative. Hence, we propose plotting other extreme scenarios, as shown in Figure 4, where we consider different values of partial $R^2$ with the outcome, including 75% and 50%.

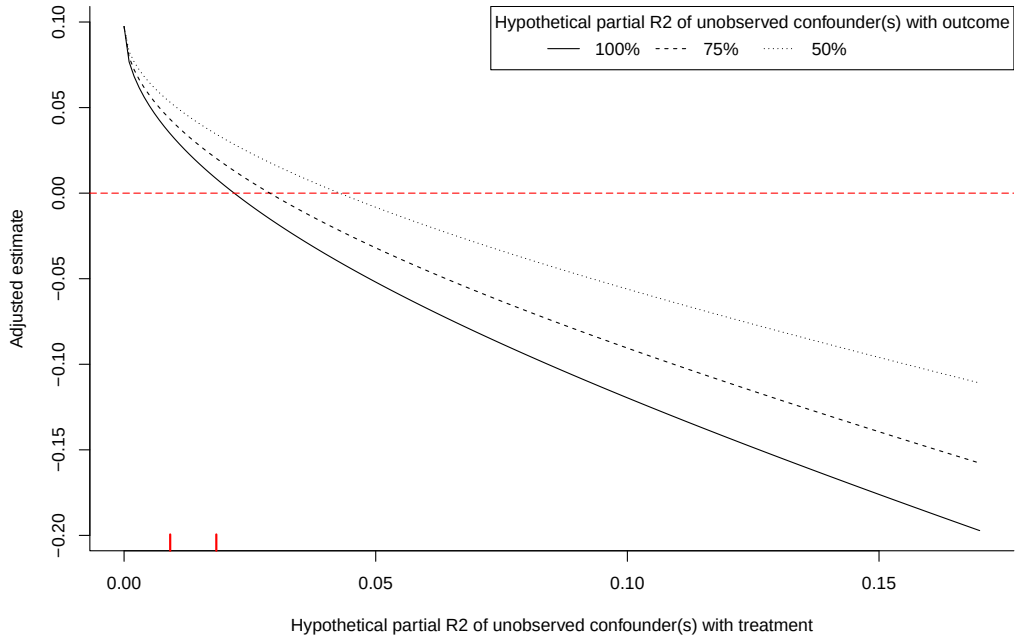


Figure 4: Extreme scenarios sensitivity analysis

**Note**: Extreme scenarios sensitivity analysis using partial $R^2$ including benchmark bounds using observed covariates. The horizontal axis shows the partial $R^2$ of the confounder(s) with the treatment and the vertical axis the adjusted treatment effect estimate. The partial $R^2$ of the confounder with the outcome is represented by *different curves* for each scenario. The most extreme scenario shown on the plot is the solid curve, in which the unobserved confounders explain *all residual variance* of the outcome. In this extreme-scenario, in our running example, the confounder(s) would have to explain 2.2% of the residual variance of the treament to bring down the estimated effect to zero (which is exactly the partial $R^2$ of the treatment with the outcome). Note this is below the two bounds of a confounder one time or twice as strong as *Female*, shown in the red benchmark lines.

# 6 Discussion

## 6.1 Making formal sensitivity analysis standard practice

Given that ruling out unobserved confounders is often difficult or impossible in observational research, one might expect that sensitivity analyses would be a routine procedure in numerous disciplines. Why then are they not commonplace? We surmise there are three main obstacles.

First, the assumptions that many methods impose on the nature and distribution of unobserved confounders may be difficult to sustain in some cases. Rosenbaum and Rubin (1983a), Imbens (2003) and Dorie et al. (2016) require specifying the distribution of the confounder as well as modeling the treatment assignment mechanism; the methods put forward in Heckman et al. (1998), Robins (1999), Brumback et al. (2004) and Blackwell (2013) need to directly specify a confounding function relating selection to potential outcomes. Assessing the sensitivity to *some* form of confounding is surely an improvement over simply assuming no confounding. However, widespread adoption of sensitivity analysis would benefit from methods that weaken these assumptions. Our derivations are rooted in traditional the OVB precisely to avoid simplifying assumptions about the unobserved confounder. As we have seen, the partial $R^2$ parameterization allows a flexible sensitivity analysis framework for assessing the sensitivity to multiple (possibly nonlinear) confounders, even including mispecification of the functional form of observed covariates.

A related approach, which also hinges on OVB for sensitivity to unobserved confounders, is Hosman et al. (2010). Their method, however, suffers from two main deficiencies which we address in this paper. The first is the use of informal benchmarking procedures to make claims about the sensitivity to confounders "not unlike" included variables in their relationship to the treatment and outcome. As discussed, such comparisons can be seriously misleading. A second problem is the choice of parameterization: Hosman et al. (2010) ask researchers to calibrate intuitions about the strength of the confounder with the treatment using a t-value. This is a problematic choice, both because the t-value has little substantive meaning, and relatedly, because it incorporates information on both the strength of association and the sample size, the later being irrelevant for identification concerns.[15] By contrast, the partial $R^2$ is directly relevant to identification, and has a clear substantive interpretation as the proportion of variance explained.

This leads to a second obstacle to wider adoption of sensitivity analysis: the lack of simple, interpretable measures users can report alongside other regression summary statistics. Our minimal reporting recommendation for regression tables (see Table 1) aims to fill this gap with: (i) the robustness value, indicating the minimal strength of equal association a confounder needs to have to change the research conclusions, and (ii) the $R^2_{Y\sim D|\boldsymbol{X}}$, which works as an extreme-scenario sensitivity analysis. Regarding the robustness value in particular, a related proposal suited for the risk ratio has recently been advocated by VanderWeele and Ding (2017), which they called the E-value. The robustness value parameterizes the association of the confounder with the treatment and the outcome in terms of percentage of variance explained (the partial $R^2$), whereas the E-value parameterizes these in terms of risk ratios. Whether one scale is preferable over the other depends on context, and researchers should be aware of both options. For other effect measures, such as risk differences,

[15] The t-value on the expression of the bias is an artifact of both multiplying and dividing by the degrees of freedom, as in our Equation 12. While t-statistics can be useful in sensitivity analysis for computational purposes (to utilize quantities routinely reported in regression tables), their dependence on sample size makes them an akward choice for contemplating how strongly related a confounder is to the treatment. Consider a t-value of 200. With 100 degrees of freedom, the confounder explains virtually all the residual variance of the treatment (partial $R^2$ of 0.9975), while with 10 million degrees of freedom, the confounder explains less than 0.5%. These are clearly very different confounders, and the partial $R^2$ makes this distinction upfront.

the E-value is approximate, whereas if the researcher uses linear regression for such estimate, the robustness value is exact. Overall, we believe the dissemination of measures such as the E-value and the robustness value is an important step towards the widespread adoption of sensitivity analysis to unobserved confounding. In current practice, robustness is often informally or implicitly linked to t-values or p-values, neither of which measure how sensitive an estimate is to unobserved confounding. The extension of the robustness value to non-linear models is worth exploring in future research.

Finally, a third and fundamental obstacle to the use of sensitivity is the difficulty in connecting the results of a formal sensitivity analysis to the researcher's substantive knowledge. This can be only partially overcome by statistical tools, as it relies upon the nature of expert knowledge used for plausibility judgments. As we have discussed, prior work has suggested benchmarking procedures informally using observed statistics of observables (Imbens 2003; Hosman et al. 2010; Blackwell 2013; Dorie et al. 2016; Middleton et al. 2016), a practice which leads to undesirable consequences, as shown in Appendix A.2. We show here instead how one can formally bound the strength of the unobserved confounders as strong (or multiple times stronger) than observed covariates. When researchers can credibly argue to have measured the most important determinants of the treatment assignment and of the outcome, such bounding could be a valuable tool.

Another formal bounding argument has been also presented in Oster (2017). Unlike the informal benchmarking practices previously discussed, Oster (2017) bounding procedure is formally correct. However, the parameterization of the procedure makes meaningful reasoning about these parameters very difficult in a real research context. More precisely, Oster (2017) asks researchers to make plausibility judgments on two sensitivity parameters, $R_{\max}$ and $\delta_{\text{Oster}}$. The $R_{\max}$ parameter has a simple and direct interpretation—it is the maximum explanatory power that one could have with the full outcome regression, i.e., $R_{\max} = R^2_{Y\sim D+\boldsymbol{X}+\boldsymbol{Z}}$. This quantity has a one to one relationship with $R^2_{Y\sim \boldsymbol{Z}|\boldsymbol{X},D}$

$$R^2_{Y\sim \boldsymbol{Z}|\boldsymbol{X},D} = \frac{R_{\max} - R^2_{Y\sim D+\boldsymbol{X}}}{1 - R^2_{Y\sim D+\boldsymbol{X}}} \tag{24}$$

However, the second sensitivity parameter, $\delta_{\text{Oster}}$, is not easily interpretable in substantive terms. Define indexes $W_1 := \boldsymbol{X}\hat{\beta}$ and $W_2 := \boldsymbol{Z}\hat{\gamma}$, where $\boldsymbol{X}$ is a matrix of observed covariates and $\boldsymbol{Z}$ a matrix of unobserved covariates. Critically, $\hat{\beta}$ and $\hat{\gamma}$ are chosen such that $Y = \hat{\tau}D + W_1 + W_2 + \hat{\varepsilon}_{\text{full}}$.[16] Then, $\delta_{\text{Oster}} = \text{cov}(W_2, D)/\text{var}(W2) \times \text{var}(W_1)/\text{cov}(W_1, D)$. Oster (2017) regards this parameter as a measure of "proportional selection", i.e. how strongly the unobservables drive treatment assignment relative to the observables. However, $\delta_{Oster}$ captures not only the relative influence of $\boldsymbol{X}$ and $\boldsymbol{Z}$ over the treatment, but also their association with the outcome because $W_1$ and $W_2$ have been defined through association with the outcome. To examine the simple case with only one covariate and one confounder and assuming $X \perp Z$, we have

$$\delta_{\text{Oster}} = \frac{\text{cov}(W_2, D)}{\text{var}(W_2)} \frac{\text{var}(W_1)}{\text{cov}(W_1, D)} = \frac{\text{cov}(\hat{\gamma}Z, D)}{\text{var}(\hat{\gamma}Z)} \frac{\text{var}(\hat{\beta}X)}{\text{cov}(\hat{\beta}X, D)} = \frac{\text{cov}(Z, D)}{\hat{\gamma}\text{var}(Z)} \frac{\hat{\beta}\text{var}(X)}{\text{cov}(X, D)} = \frac{\hat{\lambda}}{\hat{\gamma}} \frac{\hat{\beta}}{\hat{\theta}} \tag{25}$$

where $\hat{\lambda}$ and $\hat{\theta}$ are the coefficients of the regression, $D = \hat{\theta}X + \hat{\lambda}Z + \hat{\varepsilon}_D$. Claims that $\delta_{\text{Oster}} = 1$ implies "the unobservable and observables are equally related to the treatment" (Oster 2017, p.6) can lead researchers astray, as this quantity also depends upon associations with the outcome.

[16] Oster (2017) uses population values. Here we will use sample values to maintain consistency with the rest of the paper, but this has no consequence for the argument in question.

For example, let the variables be standardized to unit variance and pick $\hat{\beta} = \hat{\theta} = p$ , $\hat{\gamma} = \hat{\lambda} = p/2$. This is a case where the confounder $Z$ has half the explanatory power of $X$ (as measured by standardized coefficients), yet $\delta_{\text{Oster}} = 1$. While researchers may be able to make arguments about relative explanatory power of observables and unobservables in the treatment assignment process, the $\delta_{Oster}$ parameter does not correspond directly to such claims.[17] By contrast, the parameter $k_D$ we introduce in our bounding procedure (Section 4.4) directly captures the relative explanatory power of the unobservable and observable over treatment assignment, in terms of partial $R^2$ (or total $R^2$, depending on the investigator's preference).

Such parameterization choices are not innocuous when they drive a wedge between what investigators can argue about and the values of the parameters these arguments imply. It is thus important that the sensitivity parameters used in these exercises be as transparent as possible and match investigators' conception of what the parameters imply. Hence, we employ $R^2$ based parameters, rather than t-values or quantities relating indexes. The resulting sensitivity parameters not only correspond more directly to what investigators can articulate and reason about, but also lead to the rich set of sensitivity exercises we have discussed. Of course, further improvements may be possible and future research should investigate whether such flexibility can be achieved with yet more meaningful parameterizations. The tools we propose here, like any other, have potential for abuse. We thus end with important caveats, in particular emphasizing that sensitivity analysis should not be used for automatic judgment, but as an instrument for disciplined arguments about confounding.

## 6.2 Sensitivity analysis as principled argument

Sensitivity analyses tell us what we would have to be prepared to believe in order to accept the substantive claim initially made (Rosenbaum 2005). The sensitivity exercises proposed here tell the researcher how strong unobserved confounding would have to be in order to meaningfully change the treatment effect estimate beyond some level we are interested in, and employ observed covariates to argue for bounds on such confounding where possible. Whether we can rule out such confounders depends on expert judgment. As a consequence, the research design, identification strategy as well as the story explaining the quality of the covariates used for benchmarking all play central roles.

We strongly warn against blindly employing covariates for bounding the strength of confounders, without the ability to argue that the covariates are likely to be among the strongest predictors of the outcome or treatment assignment. A particular moral hazard here is that weak covariates can make the apparent bounds look better. It is thus important for readers and reviewers to demand that researchers properly justify and interpret their sensitivity results, after which such claims can be debated.

We also do not propose arbitrary threshold values for the sensitivity statistics described here, such as the RV, the $R^2_{Y\sim D|\boldsymbol{X}}$, or the strength of relationship of $Z$ to $Y$ or $D$. In our view, no meaningful universal threshold is possible to establish. The strength of confounding an investigator can rule out depends critically upon the research design, special knowledge of the situation and what unobservables may be involved in treatment assignment or the outcome, and the quality of the observed covariates in this regard. For instance, in a poorly controlled regression on observational

[17] Indeed, arguments made by researchers applying Oster (2017) suggest they believe they are comparing the explanatory power of observables and unobservables over treatment assignment in terms such as correlation or variance explained. e.g. "Following the approach suggested by Altonji, Elder, and Taber (2005) and Oster (2017), we estimate that unobservable country-level characteristics would need to be 1.44 times more correlated with treatment than observed covariates to fully explain the apparent impact of grammatical gender on the level of female labor force participation; unobserved factors would need to be 3.23 times more closely linked to treatment to explain the impact of grammatical gender on the gender gap in labor force participation." (Jakiela and Ozier 2018, p.4)

data with no clear understanding of what (unobservables) might influence treatment uptake, it would be very difficult to credibly claim that a RV of 5% is "good news". On the other hand, in a quasi-experiment where the treatment is assigned in such a way that certain observed covariates account for almost any possible selection, then a more credible case can be made that the types of confounders that would substantially alter our conclusions are unlikely. Sensitivity analyses is best suited as a tool for disciplined discussion, not a tool to dismiss such discussion by following automatic procedures.[18]

A final point of concern is publication bias. Sensitivity analysis should not be misappropriated as a tool for inhibiting "imperfectly identified" research on important topics. Important questions, using state-of-the-art research design, which turns out to not be robust to reasonable sources of confounding should not be dismissed. On the contrary, with sensitivity analyses, instead, we can conduct an imperfect investigation, while transparently revealing how susceptible our results are to confounding. This gives future researchers a starting point and roadmap for improving upon the robustness of these answers in their following inquiries.

[18] Accordingly, we also recommend against current practices using the suggestion of Oster (2017) to take $\delta_{\text{Oster}} \leq 1$ as an appropriate bound in very different contexts where the "room for confounding" may vary enormously.

## References


Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J. D. and Pischke, J.-S. (2017). Undergraduate econometrics instruction: Through our classes, darkly. Technical report, National Bureau of Economic Research.

Blackwell, M. (2013). A selection bias approach to sensitivity analysis for causal effects. *Political Analysis*, 22(2):169–182.

Brumback, B. A., Hernán, M. A., Haneuse, S. J., and Robins, J. M. (2004). Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in medicine*, 23(5):749–767.

Chen, B. and Pearl, J. (2015). Exogeneity and robustness. Technical report, Tech. Rep.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *journal of National Cancer Institute*, (23):173–203.

Ding, P. and Miratrix, L. W. (2015). To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *Journal of Causal Inference*, 3(1):41–57.

Dorie, V., Harada, M., Carnegie, N. B., and Hill, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in medicine*, 35(20):3453–3470.

Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

Flint, J. and de Waal, A. (2008). *Darfur: a new history of a long war*. Zed Books.

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, 29(2):147–194.

Frank, K. A., Maroulis, S. J., Duong, M. Q., and Kelcey, B. M. (2013). What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis*, 35(4):437–460.

Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401.

Hazlett, C. (2013). Angry or weary? The effect of personal violence on attitudes towards peace in darfur. *Working Paper. Available at http://www.mit.edu/∼hazlett*.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. Technical report, National bureau of economic research.

Hosman, C. A., Hansen, B. B., and Holland, P. W. (2010). The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *The Annals of Applied Statistics*, pages 849–870.

Human Rights Watch (2004). Darfur destroyed: Ethnic cleansing by government and militia forces in western sudan.

Imai, K., Keele, L., Yamamoto, T., et al. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, 25(1):51–71.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, 93(2):126–132.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jakiela, P. and Ozier, O. (2018). Gendered language. *Policy Research Working Paper, World Bank*.

Leamer, E. E. (2016). S-values: Conventional context-minimal measures of the sturdiness of regression coefficients. *Journal of Econometrics*, 193(1):147 – 161.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010.

Lovell, M. C. (2008). A simple proof of the FWL theorem. *The Journal of Economic Education*, 39(1):88–91.

Middleton, J. A., Scott, M. A., Diakow, R., and Hill, J. L. (2016). Bias amplification and bias unmasking. *Political Analysis*, 24(3):307–323.

Oster, E. (2014). Unobservable selection and coefficient stability: Theory and evidence. *NBER working paper*.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pages 1–18.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. (2011). Invited commentary: understanding bias amplification. *American journal of epidemiology*, 174(11):1223–1227.

Pearl, J. (2012). On a class of bias-amplifying variables that endanger effect estimates. *arXiv preprint arXiv:1203.3503*.

Pearl, J. (2015). Comment on Ding and Miratrix: To adjust or not to adjust? *Journal of Causal Inference*, 3(1):59–60.

Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese*, 121(1):151–179.

Rosenbaum, P. R. (2002). Observational studies. In *Observational studies*, pages 1–17. Springer.

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In *Encyclopedia of statistics in behavioral science*, volume 4, pages 1809, 1814. John Wiley & Sons Ltd.

Rosenbaum, P. R. and Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 212–218.

Rosenbaum, P. R. and Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Vanderweele, T. J. and Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology (Cambridge, Mass.)*, 22(1):42–52.

VanderWeele, T. J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the e-value. *Annals of Internal Medicine*, 167(4):268–274.

# A Appendices [Draft]

## A.1 Simple measures for routine reporting

### Preliminaries

For any univariate regression, recall $R^2 = t^2/(t^2+\text{df})$, $t^2 = \left(\frac{R^2}{1-R^2}\right)\text{df}$, and $f^2 = \frac{R^2}{1-R^2} = \frac{t^2}{\text{df}}$, where df is the regression's degrees of freedom. Repeating the partialing out procedure to alow for covariates, the partial $R^2$ of any covariate can be written in terms of its coefficient's $t$ statistic and *vice-versa*.

For instance, the partial $R^2$ of the confounder with the treatment, conditional on $\boldsymbol{X}$, can be written as

$$R^2_{D\sim Z|\boldsymbol{X}} = \frac{t^2_{\hat{\delta}}}{t^2_{\hat{\delta}} + \text{df}} \tag{26}$$

analogously,

$$f^2_{D\sim Z|\boldsymbol{X}} = \frac{R^2_{D\sim Z|\boldsymbol{X}}}{1 - R^2_{D\sim Z|\boldsymbol{X}}} = \frac{t^2_{\hat{\delta}}}{\text{df}} \tag{27}$$

### General strength of a confounder

Consider a confounder strong enough to change the estimated treatment effect by $(100 \times q)\%$. This means that $|\widehat{\text{bias}}| = q|\hat{\tau}_{\text{res}}|$. Hence, by equation 13 we have that:

$$q|\hat{\tau}_{\text{res}}| = \sqrt{\frac{R^2_{Y\sim Z|\boldsymbol{X},D}\ R^2_{D\sim Z|\boldsymbol{X}}}{1 - R^2_{D\sim Z|\boldsymbol{X}}}}\text{se}(\hat{\tau}_{\text{res}})\sqrt{\text{df}} \tag{28}$$

Since $\frac{|\hat{\tau}_{\text{res}}|}{\text{se}(\hat{\tau}_{\text{res}})\sqrt{\text{df}}} = \frac{|t_{\hat{\tau}_{\text{res}}}|}{\sqrt{\text{df}}} = f_{Y\sim D|\boldsymbol{X}}$ dividing both sides by $\text{se}(\hat{\tau}_{\text{res}})\sqrt{df}$, we obtain:

$$q|f_{Y\sim D|\boldsymbol{X}}| = \sqrt{\frac{R^2_{Y\sim Z|\boldsymbol{X},D}\ R^2_{D\sim Z|\boldsymbol{X}}}{1 - R^2_{D\sim Z|\boldsymbol{X}}}} \tag{29}$$

$$= |R_{Y\sim Z|\boldsymbol{X},D} \times f_{D\sim Z|\boldsymbol{X}}| \tag{30}$$

That is, to bring the estimated effect down by $(100 \times q)\%$, the bias factor of the confounder $\left(R_{Y\sim Z|\boldsymbol{X},D} f_{D\sim Z|\boldsymbol{X}}\right)$ has to equal $q$ times the partial $f$ of the treatment with the outcome.

### Extreme sensitivity scenario with $R^2_{Y\sim D|\boldsymbol{X}}$

Considering the extreme case scenario where the confounders explain all the residual variance of the outcome, that is, $R^2_{Y\sim Z|\boldsymbol{X},D} = 1$, a confounder strong enough to bring down the estimated effect to zero, that is $q = 1$, would need to satisfy $f^2_{Y\sim D|\boldsymbol{X}} = f^2_{D\sim Z|\boldsymbol{X}}$ which implies $R^2_{Y\sim D|\boldsymbol{X}} = R^2_{D\sim Z|\boldsymbol{X}}$. This shows the partial $R^2$ of the treatment with the outcome is itself a measure of an extreme-scenario sensitivity analysis.

**The Robustness Value (RV)**

Now consider a confounder with $R^2_{Y\sim Z|\boldsymbol{X},D} = R^2_{D\sim Z|\boldsymbol{X}} = RV_q$. Rearranging terms and squaring Equation 29, one obtains

$$RV_q^2 + f_q^2 RV_q - f_q^2 = 0 \tag{31}$$

Where $f_q := q|f_{Y\sim D|\boldsymbol{X}}|$. Solving the quadratic equation for $RV_q$,

$$RV_q = \frac{1}{2}\left(\sqrt{f_q^4 + 4f_q^2} - f_q^2\right) \tag{32}$$

gives us the equation for the robustness value for the point estimate.

**RV for t-values, or lower and upper bounds of confidence intervals**

Imagine now the researcher wants to know how strong a confounder would need to be for a $\alpha\%$ confidence interval to include a change of $(100 \times q)\%$ of the treatment estimate. Consider again a confounder with equal association with the treatment and the outcome, $R^2_{Y\sim Z|\boldsymbol{X},D} = R^2_{D\sim Z|\boldsymbol{X}} = RV_{q,\alpha}$. By Equation 13,

$$|\hat{\tau}| = |\hat{\tau}_{\text{res}}| - \text{se}(\hat{\tau}_{\text{res}})\frac{RV_{q,\alpha}}{\sqrt{1 - RV_{q,\alpha}}}\sqrt{df} \tag{33}$$

where we are assuming the bias reduces the absolute value of the estimated effect. For the opposite direction just change the subtraction to addition. We further have that Equation 12 simplifies to,

$$\text{se}(\hat{\tau}) = \text{se}(\hat{\tau}_{\text{res}})\sqrt{\frac{df}{df-1}} \tag{34}$$

Let $|t^{\alpha}_{df-1}| \leq q|f_{Y\sim D|\boldsymbol{X}}|$ denote the t-value threshold for a t-test of significance level $\alpha$ and $df - 1$ degrees of freedom. For the adjusted t-test to not reject the hypothesis $H_0 : \tau = (1-q)|\hat{\tau}_{\text{res}}|$, we must have that:

$$|t^{\alpha}_{df-1}| = \frac{|\hat{\tau}| - (1-q)|\hat{\tau}_{\text{res}}|}{\text{se}(\hat{\tau})} \tag{35}$$

$$= \frac{q|\hat{\tau}_{\text{res}}| - \text{se}(\hat{\tau}_{\text{res}})\frac{RV_{q,\alpha}}{\sqrt{1-RV_{q,\alpha}}}\sqrt{df}}{\text{se}(\hat{\tau}_{\text{res}})\sqrt{\frac{df}{df-1}}} \tag{36}$$

$$= \left(q|f_{Y\sim D|\boldsymbol{X}}| - \frac{RV_{q,\alpha}}{\sqrt{1 - RV_{q,\alpha}}}\right)\sqrt{df-1} \tag{37}$$

Now rearrange terms and square it to obtain,

$$RV_{q,\alpha}^2 + f_{q,\alpha}^2 RV_{q,\alpha} - f_{q,\alpha}^2 = 0 \tag{38}$$

Where,

$$f_{q,\alpha} := q|f_{Y\sim D|\boldsymbol{X}}| - \frac{|t_{df-1}^{\alpha}|}{\sqrt{df-1}} \tag{39}$$

Solving the quadratic equation for $RV_{q,\alpha}$ we obtain the robustness value for a reduction of $(100 \times q)\%$ to not be rejected at the significance level $\alpha$:

$$RV_{q,\alpha} = \frac{1}{2}\left(\sqrt{f_{q,\alpha}^4 + 4f_{q,\alpha}^2} - f_{q,\alpha}^2\right) \tag{40}$$

Notice that if one picks $|t_{df-1}^{\alpha}| = 0$ then $RV_{q,\alpha}$ reduces to $RV_q$. Also note that, for fixed $|t_{df-1}^{\alpha}|$, when $df \to \infty$ we have that $RV_{q,\alpha} \to RV_q$.

## A.2 Problems with "naive" benchmarking

Many researchers have suggested informal benchmarking procedures using statistics of observed covariates (Imbens 2003; Hosman et al. 2010; Blackwell 2013; Dorie et al. 2016; Middleton et al. 2016). This practice has undesirable properties, because the observed statistics used as benchmarks are themselves affected by the omission of the confounder. Thus, claims of the type, "a confounder $Z$ not unlike $X$ could not change the research conclusions", using those informal benchmarks can be seriously misleading.

Consider for a moment the difference between the coefficient on $\boldsymbol{X}$ in the full Equation 3, and its estimate in the restricted Equation 4, $\hat{\beta}_{\text{res}}$ and $\hat{\beta}$. Using the same OVB approach of "impact times imbalance", we arrive at

$$\hat{\beta}_{\text{res}} - \hat{\beta} = \hat{\gamma}\hat{\psi} \tag{41}$$

where $\hat{\psi}$ is obtained from the regression $Z = \hat{\delta}D + \boldsymbol{X}\hat{\psi} + \hat{\varepsilon}_Z$. Note that $\hat{\psi}$ can be non-zero even if $\boldsymbol{X} \perp Z$. This happens because $D$ is a collider (Pearl 2009), and conditioning on $D$ creates dependency between $Z$ and $\boldsymbol{X}$. Therefore, claims comparing $\hat{\beta}_{\text{res}}$ with $\hat{\gamma}$ as if $\hat{\beta}_{\text{res}}$ were a good proxy for $\hat{\beta}$ can be very misleading. This reasoning holds whether one is using the regression coefficients themselves or other observed statistics, such as partial $R^2$ values or t-values.

To illustrate this threat more concretely, consider a simple simulation based on the structural model:

$$Y = X + Z + \varepsilon_y \tag{42}$$
$$D = X + Z + \varepsilon_d \tag{43}$$
$$X = \varepsilon_x \tag{44}$$
$$Z = \varepsilon_z \tag{45}$$

where all disturbances, $\varepsilon_y$, $\varepsilon_d$, $\varepsilon_x$ and $\varepsilon_z$ are independent standard normal random variables. Notice there is no effect of $D$ on $Y$, thus any observed association is due to confounding.

Simulating this data generating process with a sample of size 10,000, we obtain $\hat{\tau}_{\text{res}} = 0.5$, $\text{se}(\hat{\tau}_{\text{res}}) = 0.0027$. The unadjusted parameters one would use to employ $X$ for benchmarking would be $R^2{}_{Y\sim X|D} = 0.10$ and $R^2{}_{D\sim X} = 0.34$, resulting in a computed bias of 0.19 (Equation 13), or an adjusted effect estimate of approximately 0.31.

The results are shown in Figure 5. Note the informal benchmark point is still far away from zero, and the investigator would incorrectly conclude that a confounder "not unlike $X$" would not be sufficient to bring down the estimated effect to zero. However, this is clearly misleading. The confounder $Z$ not only is exactly like $X$ but it would also bring down the estimated effect to zero, since that is the truth. This informal benchmarking procedure can thus generate a false sense of confidence, even if the investigator correctly assumed that the confounder is "no worse" than $X$.

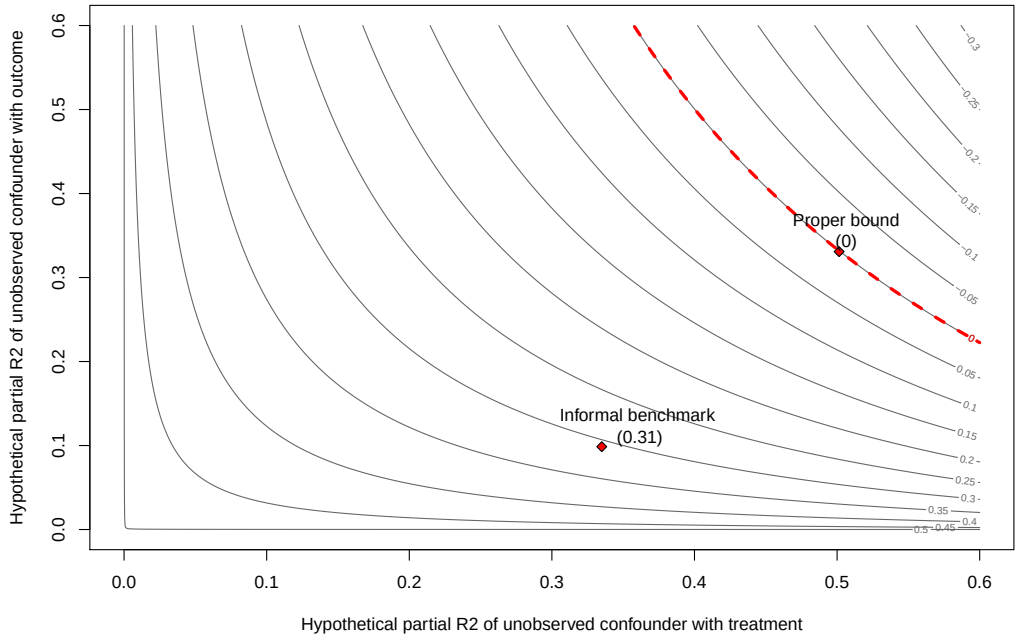


Figure 5: Sensitivity contours of point estimate — informal benchmarking *versus* proper bounds

Figure 5 also shows formal bounds obtained with the procedures given in the next section. Note these would lead the researcher to the correct conclusion: an unobserved confounder with the same strength as $X$ could be powerful enough to bring down the estimated effect to zero.

## A.3 Formal benchmark bounds

Suppose the researcher has substantive knowledge that certain covariates are "the most important predictors of the outcome" and other covariates "the most important predictors of the treatment assignment". Imagine, also, that the researcher is willing to defend the claim that the unobserved confounder $Z$ is not "as important" as those covariates, at least in some notion of multiple.

In order to use this information for bounding the strength of the confounder $Z$, we need to give it an operational meaning. We will substantiate these types of claim as comparisons of the explanatory power of the confounder vis-a-vis the explanatory power of the observed covariates. Mathematically, we can quantify these comparisons using total or partial $R^2$ measures. Here we will assume that

$Z \perp \boldsymbol{X}$ or, equivalently, that the following analysis applies to the part of $Z$ not linearly explained by covariates $\boldsymbol{X}$.

### A.3.1 Comparing total $R^2$ of covariates with total $R^2$ of confounder

Although in the text we use the bounds by comparing partial $R^2$ measures, perhaps the simplest derivation is the comparison of the total $R^2$ of observed covariates with the total $R^2$ of the unobserved confounder $Z$. To make ideas clear, let us consider an example. Consider the observed covariate $X_j$, assumed to be an important predictor of the treatment assignment $D$. If the researcher believes the correlation of $X_j$ with $D$ to be stronger than the correlation of $Z$ with $D$, this implies,

$$R^2_{D\sim Z} < R^2_{D\sim X_j} \tag{46}$$

We could use the same argument for comparing $R^2_{Y\sim Z}$ with $R^2_{Y\sim X_j}$. It turns out these types of claim are enough for bounding the sensitivity parameters. Let us generalize this notion by defining,

$$k_D := \frac{R^2_{D\sim Z}}{R^2_{D\sim X_j}}, \qquad k_Y := \frac{R^2_{Y\sim Z}}{R^2_{Y\sim X_j}} \tag{47}$$

That is, $k_D$ and $k_Y$ measure how the correlation of $Z$, with $D$ and $Y$, compares to the correlation of $X_j$ with those same variables. Our goal here is to re-express both sensitivity parameters as a function of $k_D$ and $k_Y$. Since $Z \perp \boldsymbol{X}$, we have that

$$R^2_{D\sim Z+\boldsymbol{X}} = R^2_{D\sim Z} + R^2_{D\sim \boldsymbol{X}} = k_D R^2_{D\sim X_j} + R^2_{D\sim \boldsymbol{X}} \tag{48}$$

$$R^2_{Y\sim Z+\boldsymbol{X}} = R^2_{Y\sim Z} + R^2_{Y\sim \boldsymbol{X}} = k_Y R^2_{Y\sim X_j} + R^2_{Y\sim \boldsymbol{X}} \tag{49}$$

Now we can trivially re-express $R^2_{D\sim Z|\boldsymbol{X}}$ as function of $k_D$,

$$R^2_{D\sim Z|\boldsymbol{X}} = \frac{R^2_{D\sim Z+X} - R^2_{D\sim \boldsymbol{X}}}{1 - R^2_{D\sim \boldsymbol{X}}} \tag{50}$$

$$= k_D \left( \frac{R^2_{D\sim X_j}}{1 - R^2_{D\sim \boldsymbol{X}}} \right) \tag{51}$$

Obviously, analogous result holds for $R^2_{Y\sim Z|\boldsymbol{X}}$. What is left to us is to re-express $R^2_{Y\sim Z|\boldsymbol{X},D}$. Using the standard recursive definition of partial correlations, we know that

$$\left|R_{Y\sim Z|\boldsymbol{X},D}\right| = \frac{\left|R_{Y\sim Z|\boldsymbol{X}} - R_{Y\sim D|\boldsymbol{X}} R_{D\sim Z|\boldsymbol{X}}\right|}{\sqrt{1 - R^2_{Y\sim D|\boldsymbol{X}}}\sqrt{1 - R^2_{D\sim Z|\boldsymbol{X}}}} \tag{52}$$

Notice the only two unknown terms of the RHS including the confounder, $R_{Y\sim Z|\boldsymbol{X}}$ and $R_{D\sim Z|\boldsymbol{X}}$, were already re-expressed as a function of $k_D$ and $k_Y$, as we have shown. We now show that we can determine the sign of the correlations, by considering the direction of the strengths of the confounder that act towards hurting our prefered hypothesis.

Let us assume the confounder acts towards reducing the absolute value of the effect size. If the

effect size is positive ($R_{Y\sim D|\boldsymbol{X}} > 0$), this means $R_{Y\sim Z|\boldsymbol{X},D}$ and $R_{D\sim Z|\boldsymbol{X}}$ must have the same signs. Consider, first, $R_{Y\sim Z|\boldsymbol{X},D} < 0$ and $R_{D\sim Z|\boldsymbol{X}} < 0$. This implies $R_{Y\sim Z|\boldsymbol{X}} < 0$, which means we are reducing the absolute value of $R_{Y\sim Z|\boldsymbol{X}}$. Now consider $R_{Y\sim Z|\boldsymbol{X},D} > 0$ and $R_{D\sim Z|\boldsymbol{X}} > 0$. This implies $R_{Y\sim Z|\boldsymbol{X}} > 0$, which, again, means we are reducing the absolute value of $R_{Y\sim Z|\boldsymbol{X}}$. If the effect size is negative ($R_{Y\sim D|\boldsymbol{X}} < 0$), this now would mean that $R_{Y\sim Z|\boldsymbol{X},D}$ and $R_{D\sim Z|\boldsymbol{X}}$ must have the opposite signs, and applying the previous arguments, we reach the same conclusion that we will be reducing the absolute value of $R_{Y\sim Z|\boldsymbol{X}}$.

Therefore, considering that the confounder acts towards *reducing* the absolute value of the estimate, we have that,

$$\left|R_{Y\sim Z|\boldsymbol{X},D}\right| = \frac{|R_{Y\sim Z|\boldsymbol{X}}| - |R_{Y\sim D|\boldsymbol{X}} R_{D\sim Z|\boldsymbol{X}}|}{\sqrt{1 - R^2_{Y\sim D|\boldsymbol{X}}}\sqrt{1 - R^2_{D\sim Z|\boldsymbol{X}}}} \tag{53}$$

Finally, we point to some restrictions that the data imposes in the sensitivity parameters. First, since $0 \leq \left|R_{Y\sim Z|\boldsymbol{X},D}\right| \leq 1$, this constraints how $R_{Y\sim Z|\boldsymbol{X}}$ compares to $R_{Y\sim D|\boldsymbol{X}} R_{D\sim Z|\boldsymbol{X}}$. For instance, assuming the confounder reduces the absolute value of the effect size, we must have $|R_{Y\sim Z|\boldsymbol{X}}| \geq |R_{Y\sim D|\boldsymbol{X}} R_{D\sim Z|\boldsymbol{X}}|$. Also, since $R^2_{D\sim Z|\boldsymbol{X}} \leq 1$ we must have that $k_D \leq (1 - R^2_{D\sim \boldsymbol{X}})/R^2_{D\sim X_j}$. The same reasoning applies to $k_Y$. Extending the previous arguments to multiple covariates is straightforward, since these results hold for any subset of $\boldsymbol{X}$.

### A.3.2 Comparing partial $R^2$ of covariates with partial $R^2$ of confounders

Now imagine the researcher is willing to make a more elaborate type of claim. For instance, the researcher believes that omitting $X_j$ increases the mean squared error of the full treatment regression more than omitting $Z$. This means that, $R^2_{Y\sim \boldsymbol{X}_{-j},D,Z} < R^2_{Y\sim \boldsymbol{X},D}$, where $\boldsymbol{X}_{-j}$ represents all variables in $\boldsymbol{X}$ except $X_j$. If we then subtract $R^2_{Y\sim \boldsymbol{X}_{-j},D}$ and divide by $1 - R^2_{Y\sim \boldsymbol{X}_{-j},D}$, this gives us,

$$\frac{R^2_{Y\sim \boldsymbol{X}_{-j},D,Z} - R^2_{Y\sim \boldsymbol{X}_{-j}D}}{1 - R^2_{Y\sim \boldsymbol{X}_{-j}D}} < \frac{R^2_{Y\sim \boldsymbol{X},D} - R^2_{Y\sim \boldsymbol{X}_{-j},D}}{1 - R^2_{Y\sim \boldsymbol{X}_{-j},D}} \tag{54}$$

Which means

$$R^2_{Y\sim Z|\boldsymbol{X}_{-j},D} < R^2_{Y\sim X_j|\boldsymbol{X}_{-j},D} \tag{55}$$

That is, we can compare the strength of $Z$ to $X_j$ by assessing their relative contribution to the partial $R^2$ of the treatment regression given the remaining covariates. Generalizing this notion define,

$$k_D := \frac{R^2_{D\sim Z|\boldsymbol{X}_{-j}}}{R^2_{D\sim X_j|\boldsymbol{X}_{-j}}} \tag{56}$$

Our goal now is to re-express $R^2_{D\sim Z|\boldsymbol{X}}$ in terms of $k_D$.

**Bounding $R^2_{D\sim Z|\boldsymbol{X}}$**

From equation 56 we have that $|R_{D\sim Z|\boldsymbol{X}_{-j}}| = \sqrt{k_D}|R_{D\sim X_j|\boldsymbol{X}_{-j}}|$. Also, the assumption that $Z \perp \boldsymbol{X}$ implies $R_{Z\sim X_j|\boldsymbol{X}_{-j}} = 0$. Combining these two results, and using the standard recursive definition of partial correlations, gives us,

$$R_{D\sim Z|\boldsymbol{X}} = \left| \frac{R_{D\sim Z|\boldsymbol{X}_{-j}} - R_{D\sim X_j|\boldsymbol{X}_{-j}} R_{Z\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1 - R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1 - R^2_{Z\sim X_j|\boldsymbol{X}_{-j}}}} \right| \tag{57}$$

$$= \left| \frac{R_{D\sim Z|\boldsymbol{X}_{-j}}}{\sqrt{1 - R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}} \right| \tag{58}$$

$$= \frac{\sqrt{k_D}\left|R_{D\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1 - R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}} \tag{59}$$

$$= \sqrt{k_D}\left|f_{D\sim X_j|\boldsymbol{X}-j}\right| \tag{60}$$

Hence,

$$R^2_{D\sim Z|\boldsymbol{X}} = k_D \times f^2_{D\sim X_j|\boldsymbol{X}-j} \tag{61}$$

Also, notice that, since $R^2_{D\sim Z|\boldsymbol{X}} \leq 1$ this, means $k_D$ can't vary freely and is bounded by,

$$k_D \leq \frac{1}{f^2_{D\sim X_j|\boldsymbol{X}-j}} \tag{62}$$

As an example, if a researcher has a covariate that currently explains 50% of the residual variance of the treatment assignment (implying $f^2_{D\sim X_j|\boldsymbol{X}-j} = 1$), Equation 62 reveals it's *impossible* to have an *orthogonal* unobserved confounder $Z$ stronger than that covariate.

**Using multiple covariates** Now let us generalize the previous bound to multiple covariates. Let this set of covariates be $\mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})} = \{X_1, \ldots, X_j\}$. We will denote the complement of this set $\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}$. Thus, $k_D$ now is defined as,

$$k_D := \frac{R^2_{D\sim Z|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}{R^2_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}} \tag{63}$$

Applying the recursive definition of partial correlation to, $R_{D\sim Z|\boldsymbol{X}}$, $R_{D\sim Z|\boldsymbol{X}_{-(1)}}$, $R_{D\sim Z|\boldsymbol{X}_{-(1,2)}}$, up to $R_{D\sim Z|\boldsymbol{X}_{-(1,\ldots,j)}}$, we have that,

$$R_{D\sim Z|\boldsymbol{X}} = \frac{R_{D\sim Z|\boldsymbol{X}_{-(1,\ldots,j)}}}{\sqrt{1-R^2_{D\sim X_1|\boldsymbol{X}_{-(1)}}}\sqrt{1-R^2_{D\sim X_2|\boldsymbol{X}_{-(1,2)}}}\cdots\sqrt{1-R^2_{D\sim X_j|\boldsymbol{X}_{-(1,\ldots,j)}}}} \tag{64}$$

Since, $R^2_{D\sim Z|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}} = k_D R^2_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}$, we obtain,

$$R_{D\sim Z|\boldsymbol{X}} = \frac{\sqrt{k_D}R_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}{\sqrt{1-R^2_{D\sim X_1|\boldsymbol{X}_{-(1)}}}\sqrt{1-R^2_{D\sim X_2|\boldsymbol{X}_{-(1,2)}}}\cdots\sqrt{1-R^2_{D\sim X_j|\boldsymbol{X}_{-(1,\ldots,j)}}}} \tag{65}$$

And we can simplify this further by noticing the numerator is nothing but $\sqrt{1-R^2_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}$:

$$R_{D\sim Z|\boldsymbol{X}} = \frac{\sqrt{k_D}R_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}{\sqrt{1-R^2_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}} = \sqrt{k_D}f_{D\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}} \tag{66}$$

**Bounding** $R_{Y\sim Z|\boldsymbol{X},D}$

We have two ways of bounding $R_{Y\sim Z|\boldsymbol{X},D}$, making comparisons conditional or not conditional on $D$.

**Comparisons not conditioning on** $D$ As in the previous derivation, define,

$$k_Y := \frac{R^2_{Y\sim Z|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}}{R^2_{Y\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}}} \tag{67}$$

That is, we are asking the reseacher to compare the explanatory power of the confounder against the explanatory power of $\mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}$ with respect to the outcome, conditioning on the remaining covariates $\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}$ but *not conditioning* on the treatment. Using the same recursive argument as before, we obtain,

$$R_{Y\sim Z|\boldsymbol{X}} = \sqrt{k_Y}f_{Y\sim \mathbf{X}_{(\mathbf{1}\ldots\mathbf{j})}|\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}} \tag{68}$$

We can now bound $R^2_{Y\sim Z|\boldsymbol{X},D}$ by noting again that,

$$R_{Y\sim Z|\boldsymbol{X},D} = \frac{R_{Y\sim Z|\boldsymbol{X}} - R_{Y\sim D|\boldsymbol{X}}R_{D\sim Z|\boldsymbol{X}}}{\sqrt{1-R^2_{Y\sim D|\boldsymbol{X}}}\sqrt{1-R^2_{D\sim Z|\boldsymbol{X}}}} \tag{69}$$

And using the same argument as in Appendix A.3.2.

**Comparisons conditioning on** $D$ Here we have that $k_D$ is defined as before, but $k_Y$ compares the explanatory power of the confounder against the explanatory power of a covariate $X_j$ with respect to the outcome, conditioning on both the remaining covariates $\mathbf{X}_{-(\mathbf{1}\ldots\mathbf{j})}$ and the treatment, that is,

$$k_D := \frac{R^2_{D\sim Z|\boldsymbol{X}_{-j}}}{R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}, \qquad k_Y := \frac{R^2_{Y\sim Z|\boldsymbol{X}_{-j},D}}{R^2_{Y\sim X_j|\boldsymbol{X}_{-j},D}} \tag{70}$$

To bound $R^2_{Y\sim Z|\boldsymbol{X},D}$, we first need to investigate $R_{Z\sim X_j|\boldsymbol{X}_{-j},D}$. Expanding the partial correlation gives us,

$$\left|R_{Z\sim X_j|\boldsymbol{X}_{-j},D}\right| = \left|\frac{R_{Z\sim X_j|\boldsymbol{X}_{-j}} - R_{D\sim Z|\boldsymbol{X}_{-j}} R_{D\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{D\sim Z|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}}\right| \tag{71}$$

$$= \left|\frac{R_{D\sim Z|\boldsymbol{X}_{-j}} R_{D\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-R^2_{D\sim Z|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}}\right| \tag{72}$$

$$= \left|\frac{\sqrt{k_D} R_{D\sim X_j|\boldsymbol{X}_{-j}} R_{D\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-k_D R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}\sqrt{1-R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}}\right| \tag{73}$$

$$= \left|f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} \times f_{D\sim X_j|\boldsymbol{X}-j}\right| \tag{74}$$

where, $f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j}$ is defined to be,

$$f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} := \frac{\sqrt{k_D} R_{D\sim X_j|\boldsymbol{X}_{-j}}}{\sqrt{1-k_D R^2_{D\sim X_j|\boldsymbol{X}_{-j}}}} \tag{75}$$

Combining theses results and Equation 70 we can proceed to bound $R_{Y\sim Z|\boldsymbol{X},D}$:

$$|R_{Y\sim Z|\boldsymbol{X},D}| = \left|\frac{R_{Y\sim Z|\boldsymbol{X}_{-j},D} - R_{Y\sim X_j|\boldsymbol{X}_{-j}D} R_{Z\sim X_j|\boldsymbol{X}_{-j},D}}{\sqrt{1-R^2_{Y\sim X_j|\boldsymbol{X}_{-j}D}}\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j},D}}}\right| \tag{76}$$

$$\leq \frac{\left|R_{Y\sim Z|\boldsymbol{X}_{-j},D}\right| + \left|R_{Y\sim X_j|\boldsymbol{X}_{-j}D}\right|\left|R_{Z\sim X_j|\boldsymbol{X}_{-j},D}\right|}{\sqrt{1-R^2_{Y\sim X_j|\boldsymbol{X}_{-j}D}}\sqrt{1-R^2_{Z\sim X_j|\boldsymbol{X}_{-j},D}}} \tag{77}$$

$$= \frac{\sqrt{k_Y}\left|R_{Y\sim X_j|\boldsymbol{X}_{-j}D}\right| + \left|R_{Y\sim X_j|\boldsymbol{X}_{-j}D}\right|\left|f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} \times f_{D\sim X_j|\boldsymbol{X}-j}\right|}{\sqrt{1-R^2_{Y\sim X_j|\boldsymbol{X}_{-j}D}}\sqrt{1-f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} \times f_{D\sim X_j|\boldsymbol{X}-j}}} \tag{78}$$

$$= \left(\frac{\sqrt{k_Y} + \left|f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} \times f_{D\sim X_j|\boldsymbol{X}-j}\right|}{\sqrt{1-f^2_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}-j} \times f^2_{D\sim X_j|\boldsymbol{X}-j}}}\right)\left(\frac{|R_{Y\sim X_j|\boldsymbol{X}_{-j}D}|}{\sqrt{1-R^2_{Y\sim X_j|\boldsymbol{X}_{-j}D}}}\right) \tag{79}$$

$$= \eta\left|f_{Y\sim X_j|\boldsymbol{X}_{-j},D}\right| \tag{80}$$

Hence, we have that,

$$R^2_{Y\sim Z|\boldsymbol{X},D} \leq \eta^2 f^2_{Y\sim X_j|\boldsymbol{X}_{-j},D} \tag{81}$$

where $\eta = \frac{\sqrt{k_Y}+\left|f_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}_{-j}}\times f_{D\sim X_j|\boldsymbol{X}_{-j}}\right|}{\sqrt{1-f^2_{D\sim\sqrt{k_D}X_j|\boldsymbol{X}_{-j}}\times f^2_{D\sim X_j|\boldsymbol{X}_{-j}}}}$.

It's worth emphasizing that this bound is tight. Without further assumptions, we can always create an unobserved confounder $Z$ that makes the inequality step in 77 an equality. We can extend this last bound to multiple covariates by iteratively applying the recursive definition of partial correlation.

## A.4 Sensitivity tables

**Bias Factor table**

| $R^2_{D\sim Z\|\boldsymbol{X}}/R^2_{Y\sim Z\|\boldsymbol{X},D}$ | **5%** | **10%** | **15%** | **20%** | **25%** | **30%** | **35%** | **40%** | **45%** | **50%** | **55%** | **60%** | **65%** | **70%** | **75%** | **80%** | **85%** | **90%** | **95%** | **99%** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **5%** | 0.051 | 0.073 | 0.089 | 0.103 | 0.115 | 0.126 | 0.136 | 0.145 | 0.154 | 0.162 | 0.170 | 0.178 | 0.185 | 0.192 | 0.199 | 0.205 | 0.212 | 0.218 | 0.224 | 0.228 |
| **10%** | 0.075 | 0.105 | 0.129 | 0.149 | 0.167 | 0.183 | 0.197 | 0.211 | 0.224 | 0.236 | 0.247 | 0.258 | 0.269 | 0.279 | 0.289 | 0.298 | 0.307 | 0.316 | 0.325 | 0.332 |
| **15%** | 0.094 | 0.133 | 0.163 | 0.188 | 0.210 | 0.230 | 0.249 | 0.266 | 0.282 | 0.297 | 0.312 | 0.325 | 0.339 | 0.351 | 0.364 | 0.376 | 0.387 | 0.399 | 0.409 | 0.418 |
| **20%** | 0.112 | 0.158 | 0.194 | 0.224 | 0.250 | 0.274 | 0.296 | 0.316 | 0.335 | 0.354 | 0.371 | 0.387 | 0.403 | 0.418 | 0.433 | 0.447 | 0.461 | 0.474 | 0.487 | 0.497 |
| **25%** | 0.129 | 0.183 | 0.224 | 0.258 | 0.289 | 0.316 | 0.342 | 0.365 | 0.387 | 0.408 | 0.428 | 0.447 | 0.465 | 0.483 | 0.500 | 0.516 | 0.532 | 0.548 | 0.563 | 0.574 |
| **30%** | 0.146 | 0.207 | 0.254 | 0.293 | 0.327 | 0.359 | 0.387 | 0.414 | 0.439 | 0.463 | 0.486 | 0.507 | 0.528 | 0.548 | 0.567 | 0.586 | 0.604 | 0.621 | 0.638 | 0.651 |
| **35%** | 0.164 | 0.232 | 0.284 | 0.328 | 0.367 | 0.402 | 0.434 | 0.464 | 0.492 | 0.519 | 0.544 | 0.568 | 0.592 | 0.614 | 0.635 | 0.656 | 0.677 | 0.696 | 0.715 | 0.730 |
| **40%** | 0.183 | 0.258 | 0.316 | 0.365 | 0.408 | 0.447 | 0.483 | 0.516 | 0.548 | 0.577 | 0.606 | 0.632 | 0.658 | 0.683 | 0.707 | 0.730 | 0.753 | 0.775 | 0.796 | 0.812 |
| **45%** | 0.202 | 0.286 | 0.350 | 0.405 | 0.452 | 0.495 | 0.535 | 0.572 | 0.607 | 0.640 | 0.671 | 0.701 | 0.729 | 0.757 | 0.783 | 0.809 | 0.834 | 0.858 | 0.882 | 0.900 |
| **50%** | 0.224 | 0.316 | 0.387 | 0.447 | 0.500 | 0.548 | 0.592 | 0.632 | 0.671 | 0.707 | 0.742 | 0.775 | 0.806 | 0.837 | 0.866 | 0.894 | 0.922 | 0.949 | 0.975 | 0.995 |
| **55%** | 0.247 | 0.350 | 0.428 | 0.494 | 0.553 | 0.606 | 0.654 | 0.699 | 0.742 | 0.782 | 0.820 | 0.856 | 0.891 | 0.925 | 0.957 | 0.989 | 1.019 | 1.049 | 1.078 | 1.100 |
| **60%** | 0.274 | 0.387 | 0.474 | 0.548 | 0.612 | 0.671 | 0.725 | 0.775 | 0.822 | 0.866 | 0.908 | 0.949 | 0.987 | 1.025 | 1.061 | 1.095 | 1.129 | 1.162 | 1.194 | 1.219 |
| **65%** | 0.305 | 0.431 | 0.528 | 0.609 | 0.681 | 0.746 | 0.806 | 0.862 | 0.914 | 0.964 | 1.011 | 1.056 | 1.099 | 1.140 | 1.180 | 1.219 | 1.256 | 1.293 | 1.328 | 1.356 |
| **70%** | 0.342 | 0.483 | 0.592 | 0.683 | 0.764 | 0.837 | 0.904 | 0.966 | 1.025 | 1.080 | 1.133 | 1.183 | 1.232 | 1.278 | 1.323 | 1.366 | 1.408 | 1.449 | 1.489 | 1.520 |
| **75%** | 0.387 | 0.548 | 0.671 | 0.775 | 0.866 | 0.949 | 1.025 | 1.095 | 1.162 | 1.225 | 1.285 | 1.342 | 1.396 | 1.449 | 1.500 | 1.549 | 1.597 | 1.643 | 1.688 | 1.723 |
| **80%** | 0.447 | 0.632 | 0.775 | 0.894 | 1.000 | 1.095 | 1.183 | 1.265 | 1.342 | 1.414 | 1.483 | 1.549 | 1.612 | 1.673 | 1.732 | 1.789 | 1.844 | 1.897 | 1.949 | 1.990 |
| **85%** | 0.532 | 0.753 | 0.922 | 1.065 | 1.190 | 1.304 | 1.408 | 1.506 | 1.597 | 1.683 | 1.765 | 1.844 | 1.919 | 1.992 | 2.062 | 2.129 | 2.195 | 2.258 | 2.320 | 2.369 |
| **90%** | 0.671 | 0.949 | 1.162 | 1.342 | 1.500 | 1.643 | 1.775 | 1.897 | 2.012 | 2.121 | 2.225 | 2.324 | 2.419 | 2.510 | 2.598 | 2.683 | 2.766 | 2.846 | 2.924 | 2.985 |
| **95%** | 0.975 | 1.378 | 1.688 | 1.949 | 2.179 | 2.387 | 2.579 | 2.757 | 2.924 | 3.082 | 3.233 | 3.376 | 3.514 | 3.647 | 3.775 | 3.899 | 4.019 | 4.135 | 4.249 | 4.337 |
| **99%** | 2.225 | 3.146 | 3.854 | 4.450 | 4.975 | 5.450 | 5.886 | 6.293 | 6.675 | 7.036 | 7.379 | 7.707 | 8.022 | 8.325 | 8.617 | 8.899 | 9.173 | 9.439 | 9.698 | 9.900 |

Table 2: Bias factor table

*Note*: To use the bias factor table, first compute the absolute value of the partial (Cohen's) $f$ of the treatment with the outcome. The partial $f$ can be easily obtained in most regression tables by dividing the coefficient's t-value by the square-root of the degrees of freedom, that is, $f = \frac{t}{\sqrt{df}}$. To bring down the estimated effect to zero, the bias factor has to be greater than $|f|$. In our running example, $|f| \approx 0.15$. Looking at the table, we see that the coefficient of *Directly Harmed* would be robust to a confounder with, for instance, $R^2_{D\sim Z|\boldsymbol{X}} = 5\%$ and $R^2_{Y\sim Z|\boldsymbol{X},D} = 40\%$, but would not be robust to a confounder with $R^2_{D\sim Z|\boldsymbol{X}} = 40\%$ and $R^2_{Y\sim Z|\boldsymbol{X},D} = 5\%$. To assess the robustness of the coefficient to any change of $(100 \times q)\%$, just multiply $|f|$ by $q$. Any confounder with a bias factor less than $q|f|$ cannot cause a change of $(100 \times q)\%$ in the regression coefficient.